PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G06F 15/00, 17/60, G06G 7/52

A1

(11) International Publication Number: WO 97/00483

(43) International Publication Date: 3 January 1997 (03.01.97)

(21) International Application Number: PCT/US96/10352

(22) International Filing Date: 14 June 1996 (14.06.96)

(30) Priority Data:
08/490,984 15 June 1995 (15.06.95) US

(71) Applicant: FRAUDetect, L.L.C. [US/US]; Suite 312, 10400 Eaton Place, Fairfax, VA 22030 (US).

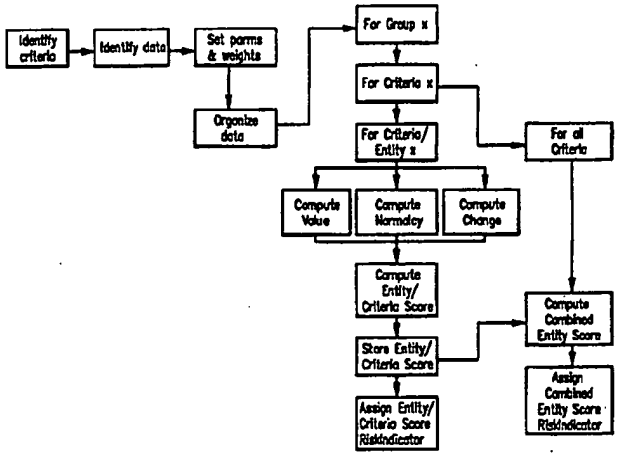
(72) Inventor: COFOD, Robert, K.; 282 Prince's Lane, Harwood, MD 20776 (US).

(74) Agent: GROSSMAN, Jon, D.; 2101 L Street, N.W., Washington, DC 20037 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: PROCESS AND APPARATUS FOR DETECTING FRAUD



(57) Abstract

A system and method for processing transaction data in filtering and targeting potentially fraudulent entities within a particular industry relies upon client data input apparatus for inputting client/provider/entity data records containing data elements for measuring an entity's performance according to pre-selected criteria, where entity date, criteria data and time data are organized and stored in three-dimensional array. A pre-processor (110) generates a score for each entity-criterion (1...n) by generating a value statistic (v), a normalcy statistic (n), and a change statistic (d), weighting each statistical value, and summing the values. An entity scoring generator (30) generates a single entity score for each entity by weighting each entity-criterion score and summing the entity-criterion scores. The entity scores are compared to pre-defined threshold values for detecting those entities most likely to be engaging in fraudulent activity and a graphical display entities according to level of risk.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

PROCESS AND APPARATUS FOR DETECTING FRAUD

FIELD OF THE INVENTION

The invention relates to a process and apparatus for analyzing client data against pre-selected criteria over a pre-determined period of time in order to identify those individuals or entities engaging in activity indicative of potential fraudulent behavior within a particular industry.

BACKGROUND OF THE ART

Many industries, such as medical and banking, suffer enormous losses due to the theft of money or services through fraudulent means, such as submitting insurance claims for unseen patients or forging checks. In the medical industry, for example, it is estimated that approximately $100 billion of U.S. annual health care expenditures is stolen through fraud or abuse. Even more alarming is that some experts believe this figure only represents the portion of which we are aware. The actual figure may be significantly higher, perhaps as much as twice the amount.

The problems surrounding fraud detection are many. Since most companies maintain relatively few investigators, limited resources are available to perform the analysis and decision-making necessary to accurately attack the problem and ensure measurable results. In addition, the possibility of wrongly accusing a client of fraudulent behavior adds an extra burden of caution and proof. Confidentiality, government involvement, market

competition, and judicial processes add other limits to the options for attacking the problem.

Other problems in fraud detection include highly varying types of targets, possible scenarios, and innovative fraud techniques requiring heavy reliance on skilled human analysts and decision makers to detect, recognize, define and initiate actions. In addition, each industry, such as the aforementioned medical and banking industries, generates a massive and dynamic quantity of information from which relevant "target" information must be extracted. Furthermore, there is a time sensitivity factor in processing and acting upon newly detected targets or instances of fraud.

Due to the above-cited difficulties in detecting those persons engaging in fraudulent activities, billions of dollars worth of stolen money or services have largely been ignored or passed on as a cost of operation.

Most cases of fraud are discovered as a result of tip-offs or indications from industry program integrity reviews. These discoveries have often been cited as being only the tip of the iceberg, yet most investigative staffs are already inundated with the cases on hand.

The problem is not unlike another well-known problem in military intelligence and targeting circles. It is known as a "target rich environment," where the number of possible targets far exceeds the resources available for attack. The military problem, like the one faced in combating health care and banking fraud and abuse, is fundamentally one of resource management. The problem

requires a method or system by which an industry may increase ensurance that the resources available to it are being spent on targets with the highest potential payback. Consequently, there is a need for processing, ordering and analyzing the volumes of transactional data, such as billing information and insurance claims, associated with a particular entity, such as a physician working within the health care industry, and extrapolating from such information whether the client is engaging in fraudulent behavior. This need for specifically targeting perpetrators of fraud or abuse becomes more readily apparent in light of the fact that there may be tens or hundreds of thousands of clients per industry. Accordingly, the present invention answers that specific need.

SUMMARY OF THE INVENTION

It is therefore apparent from the above that there exists a significant need for a system and process which effectively uses transactional data in a manner that identifies potentially fraudulent behavior and areas of greatest need for investigation. It is therefore one of the objectives of this invention to provide a fraud pre-processor detection system. Fraud can be detected for many industries dealing with a large number of clients, such as the aforementioned medical and banking industries.

Another object of this invention is to provide a process for analyzing large volumes of transaction data for the purpose of detecting indications of fraud.

Another object of this invention is to provide a system and process which automatically prioritizes those

clients or entities which are most likely to be engaging in fraudulent behavior.

Another object of this invention is to provide a system and process for establishing thresholds indicating fraud.

An additional object of this invention is to provide a system and process for filtering, processing and displaying data in a manner where it may be used as an analysis tool to support the human decision making process.

Another object of this invention is to provide a system and process for organizing data in a three dimensional array consisting of pre-selected criteria along the x-axis, groups of entities along the y-axis, and time units along the z-axis, thereby allowing entity-criterion analysis, that is analyzing each entity within a group (a single chiropractor out of a group of chiropractors) according to each criterion (e.g., the number of patients seen by the chiropractor, or the amount of claims submitted to the insurance carrier), over a pre-selected period of time units (e.g., days, months or years).

Yet a further object of this invention is to provide a system and process for generating statistical values for each entity-criterion to measure Value (v), or the relative importance of the entity with respect to the entire group, Normalcy (n), or how the entity conforms to the normal behavior of the group for a particular criterion, and Change (d), or the amount of change in activity by the entity over a period of time.

It is also an objective of this invention to weight the statistical values for Value (v), Normalcy (n) and Change (d) according to the relative importance of each value within a criterion with respect to targeting entities engaging in fraudulent activity.

Yet a further object of this invention is to add all the statistical values for each entity-criterion together to create a single entity-criterion score, weight the score according to its fraud targeting potential, sum all the individual entity-criterion scores to form a final entity score, and compare that score to a threshold value to determine the risk of fraudulent behavior.

A further object of this invention is to display all the entities under analysis in a manner that allows easy identification of those entities engaging in behavior most likely indicative of fraud.

Another object of the invention is to provide a list of actions to take against those entities suspected of fraudulent behavior so that a human analyst may select the appropriate action.

It is a further object of this invention to provide a system and process for monitoring those entities suspected of engaging in fraud once an action has been taken against the entity, and determining whether the behavior of the entity changes, and measuring such change, thereby providing some evidence that fraudulent behavior was accurately identified and corrected while also providing a mechanism by which the weighting of the relative importance of various criteria, entity-criterion

scores and entity scores can be modified to more accurately detect the presence of fraudulent behavior.

It is yet a further object of this invention to provide a system and process for generating a change meter which can be translated into dollar amounts to calculate savings to a particular industry.

Briefly described, these and other objects of the invention are accomplished by providing a system and process for inputting client transaction data for a particular industry, organizing the client transactional data into a three dimensional array, generating statistical values for each entity-criterion over time by measuring (v), (n) and (d), weighting each statistical value and summing all values to form a single entity-criterion score, weighting each entity-criterion score and summing all entity-criterion scores to form a single entity score, comparing the entity score to a pre-determined threshold, and displaying the entities so that those engaging in fraudulent behavior are easily identified.

The system and process also include providing a list of actions for the human analyst to take against those entities with scores indicating fraudulent behavior, and monitoring the behavior of those entities to see if they change thereby providing some evidence that the entities were initially engaging in fraudulent behavior, as well as providing information to more precisely adjust the weighting and threshold mechanism used by the system.

With these and other objectives, advantages and features of the invention that may become apparent, the

nature of the invention may be more clearly understood by reference to the following detailed description of the invention, the appended claims, and to the several drawings attached herein.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a hardware block diagram of the client data input apparatus of the present invention;

Fig. 2 is a hardware block diagram of the pre-processor apparatus of the present invention;

Fig. 3 is a hardware block diagram of the monitoring and tracking apparatus of the present invention;

Fig. 4 is a block flow diagram showing modules and steps for generating the Value (v) statistic;

Fig. 4A is a block flow diagram showing modules and steps for performing the sums and squares calculation used in generating the Value (v) statistic;

Fig. 5 is a block flow diagram showing modules and steps for generating the Normalcy (n) statistic;

Fig. 6 is a block flow diagram showing modules and steps for generating the Change (c) statistic;

Fig. 7 is a block flow diagram showing modules and steps for a preferred embodiment of the present invention; and

Fig. 8 is a model showing the theoretical data structure used by the present invention.

DETAILED DESCRIPTION OF THE INVENTION

Referring now to the drawings wherein like elements refer to like reference numerals throughout, Figs. 1, 2 and 3 are hardware block diagrams for a fraud pre-processor detection system according to the present invention. An objective of the present invention is to provide a mechanism for analyzing transaction data in filtering and targeting potentially fraudulent activities of individual entities. Entities may include health care providers, bank account holders, or other individuals to be analyzed for fraud risk "targeting". The term targeting refers to isolating from a universe of entities a sub-set of those entities most likely to be engaging in fraudulent activities.

Specifically, Fig. 1 illustrates the client data input apparatus for inputting and organizing large amounts of client transactional data. Fig. 2 illustrates the pre-processor apparatus for evaluating an entity by generating statistical values for each of a number of pre-selected criteria, weighting and summing each statistical value generated to form a single entity-criterion score, weighting and summing each entity-criterion score to form a single entity score, and comparing this entity score to a predetermined threshold. Fig. 3 illustrates the monitoring and tracking apparatus for displaying a "risk list" of entities with indicators as to their level of risk with respect to engaging in fraudulent activities, generating a list of actions to take against those entities with the highest risk, monitoring

those entities against which actions have been taken in order to determine if behavior has been modified, and providing data for further refining weighting and threshold values to more accurately identify fraud in light of such modified behavior.

A preferred embodiment of the data processed by this invention relates to the medical industry wherein transactional data (e.g., billing records or insurance claims) for a particular entity within a group (e.g., a cardio-thoracic surgeon within a group of surgeons or doctors, or a periodontist within a group of dentists, and so on) is analyzed to determine if patterns of fraudulent behavior can be detected. However, the processes and apparatus described herein can be applied to any other client transactional data for any number of industries.

It has been found in the preferred embodiment, that by examining client transactional data according to certain criteria, patterns can be detected. In the health care administration process, data collection occurs in a variety of ways: claims processing, subscriber enrollment, provider certification, reimbursement-accounting, and investigations. In some cases, the data is collected from external sources (e.g., provider certification). In other cases, the data is derived from internal processes (e.g., reimbursement-accounting). In health care, most of the claims/subscriber/provider related data occurs as a normal course of doing business. Also included as "data" is that information developed as a result of internal or external investigation regarding a particular case.

Administrators of large plans may involve millions of subscribers and hundreds of thousands of providers. Consequently, there is a tremendous amount of data, and this aids the perpetrator of fraud or abuse since, with some care, he or she can lie buried and virtually undetected. Because of the economics of the industry, most data processing systems are usually optimized for the rapid processing of claims, thereby limiting the kind of ad hoc analysis that is helpful in combating fraud and abuse.

As a result of the huge amount of client transactional data available for analysis, it is necessary to store in the criteria memory 16 (Fig. 1) a number of analysis criteria for use in the system as indicators of fraud. It has been found that for any domain environment subject to fraud there are many observable criteria that can be found in transaction data and history records. By understanding the operational characteristics of fraud in a particular domain, it is possible to select specific criteria than can be used to detect the risk of fraud. These criteria are determined manually through the user's knowledge of the fraud environment (e.g., how physicians commit fraud) and how the related behavior is represented in the transaction data and history. The objective of selecting the criteria is to identify independent variables that can reflect behaviors that are focused on maximizing income and minimizing investment.

In general, these criteria can include two types: (1) information which can be summed; or (2) information which can be counted. In the medical industry, examples of these criteria can include the number of claims filed by a provider (counted), the total amount billed (summed), the

number of critical versus non-critical patients (counted), the number of given patients seen by a doctor over the period of a day, week or year (counted), and so forth. These criteria can also include specific patterns indicating fraud such as billing charges for intensive care ambulance services for a non-critical patient. These specific patterns, however, address only a small number of the possible ways fraud or abuse can be committed, and are therefore of limited use. Consequently, the more general criteria lending themselves to broader analysis are preferred, such as those listed previously.

Typically, these criteria are selected according to their potential for detecting occurrences of fraud or tracking the specific behavior of an entity under analysis, usually by the standards for a particular industry as established by past experience, or by an expert in intelligence gathering and analysis. In essence, the selection of these criteria acts as one step in the filtering process whereby essential data from the large amounts of client transactional data is identified for use in the present invention. This criteria selection is analogous to a method used in military intelligence known as "indications and warning" that monitored a variety of criteria for indications of increasing threat. When such a threat was detected, an alarm was triggered to warn the analyst. This same technique is used in the preferred embodiment to scan through the behavior history of all providers or patients and identify those having the highest risk for fraud or abuse.

In the preferred embodiment, these analysis criteria are pre-programmed into the system. The analysis

criteria is used to identify the data elements contained in the data derived from the transaction process data structure received from the client data source 12. The selected data elements are included in a periodic extract process (e.g., monthly) which produces an extract data file to be processed by the pre-processor 110.

As shown in Fig. 1, once the criteria are established and programmed into the criteria memory 16, the client data converter 14 must then convert the client transaction data received from the client data source 12 for use in the pre-processor 110. Specifically, the client data converter 14 may be implemented as a hardware device or software program (such as a commercial database engine) which accepts the extract data file and transforms it physically and organizationally in preparation for its use by the pre-processor 110.

The extract data file, typically an American Standard Computer Information Interface (ASCII) file or fixed record length file, is converted to a pre-processor input data file using the data structure required by the pre-processor 110. Some of the data elements received from the client data source 12 are directly used as criteria, others must be converted before they are used. For instance, in the preferred embodiment, a criterion might be the number of claims submitted as a direct input data set. Another criterion might also use the ratio of claims per patient as an input set after dividing the number of claims submitted by the number of patients treated.

In addition to converting the data physically, the client data converter 14 also organizes the data in the

order expected by the pre-processor 110. Since peer-level comparison is one of the concepts used by the pre-processor 110, the data must be arranged into groups of like entities. In the preferred embodiment, this would mean sorting the data into physician specialty groups, in the case of banking, this would involve sorting by like account/branch combinations. The objective is to put the data into groups of entities that can be expected to exhibit common behavior based upon their functional characteristics (e.g., all cardiologists treat heart disease).

The data elements selected according to the analysis criteria can be arranged in any order except for the master criteria. The master criteria is the criteria that will have the highest consistency for reflecting an entity's performance. For instance, good master criteria for health care would be either the number of claims submitted, or the number of patients treated -- since either is required for the physician to have any activity during the month. The number of referrals, for instance, would be a poor master criteria since referrals are not required for the physician to operate.

Within each group, entity data is further organized according to time units. The pre-processor 110 can accept various time units (e.g., days, months, years, etc.) based upon the operational characteristics of the domain and the dynamics of the fraud threat. In the preferred embodiment, it is usually adequate to analyze a one year historical sample of data composed of twelve monthly time units. In the case of bank fraud, since the dynamics are much

greater, the system would analyze ninety, one day time units.

The organization of the pre-processor input data file can be theoretically envisioned as a three-dimensional cube or array (as shown in Fig. 8,) where the x-axis contains the entities under analysis, the y-axis contains the criteria that are to be measured in determining the entity's performance, and the z-axis represents time divided into time units. Accordingly, each entity within a group can be evaluated according to each criterion over a specified amount of time. This creates the concept of an entity-criterion "bucket," wherein statistical scores or values can be generated for each entity according to each criterion.

Once the data are in a form capable of being stored in an electronic database 18, they can be connected to an appropriate controller 10 which is adapted to coordinate, control, monitor and execute commands necessary for operating the client data input apparatus 100, the pre-processor apparatus 110, as well as the monitoring and tracking apparatus 120, as shown in Figs. 1, 2 and 3, respectively. The functions performed by controller 10 are communicated to the hardware elements of Figs. 1, 2 and 3 via various control lines running from the controller 10 to its associated equipment of Figs. 1, 2 and 3. The controller 10 can comprise any appropriate microprocessor or minicomputer control device including, but not limited to, a PC-based controller or a dedicated control machine. In addition, the controller can comprise software implemented as a control program.

As shown in Fig. 2, once the pre-processor input data file is constructed and stored in the client database 18, it can be fed into the statistical analysis engine 50 of the pre-processor apparatus 110. The input file is fed to the statistical analysis engine 50 as demonstrated by the following sequence: group 1, criterion 1, entity 1...n; criterion 2, entity 1...n; criterion n, entity 1...n; group 2...n, etc.. An analysis is performed on each entity-criterion combination by determining three types of statistical measures: Value (v), Normalcy (n) and Change (d). These measures are a key component in the reliability and accuracy of identifying fraudulent behavior. They are designed to detect critical characteristics related to fraudulent behavior.

The first statistical measure is generated by the value processor 22. The value processor 22 generates a Value (v) which determines "how important" the entity is within the group. This is a sort of "damage potential" measurement where the entity could be highly suspect in behavior but have an impact in dollars so small that he or she is not worth going after. The measure is simply the average of the total (t) observations for an entity, divided by the average of the total observations for the criterion among all the entities in the group, times one hundred. That is, calculate the mean for the criterion values for an entity by: (1) determining the number of active time units (Nt); (2) finding the sum of the values for all active time units (Sty); and (3) dividing Sty by Nt to find the mean of all time units for the entity (Mty).

When the Mty for each entity is calculated these values are summed for the group (S(Mty)) and then divided by the number of entities in the group (Ne) to find the mean criteria value for the group (Mc). Each entity's Mty is then divided by Mc, and the result is multiplied by 100 to find the (v). The weight (e.g., multiplier) for the value measure for this criteria is looked up in the weight table 46 and then applied to (v) to find the weighted value score (sv). This (sv) is stored in the statistical value memory 28. (Figs. 4 and 4A give detailed block flow diagrams of the processes by which the weighted value (sv) score is generated.)

The second statistical measure is generated by the normalcy processor 24. Normalcy (n) is a measure of how much the entity conforms to the normal behavior of the group for this particular criterion. The measure is important because it is rarely possible that any particular entity can know how his or her behavior appears within the behavior of the total group.

The normalcy (n) measure is the statistical "standard score" for the entity. Calculating a standard score is relatively simple and well known in the art, and therefore any routine software implementation will suffice. Once the standard score for the entity is calculated, the weight for the normalcy measure for this criterion is looked up in the weight table 46 and then applied to (n) to find the weighted normalcy score (sn). This (sn) is stored in the statistical value memory 28. (Fig. 5 gives a detailed block flow diagram of the process by which the weighted normalcy (sn) score is generated.)

The third statistical measures is generated by the change processor 26. Change (d) is measured by performing a standard single linear regression on the observation (t) values for the entity, estimating the value for the last t, comparing the estimated to the actual value for the last t, multiplying the difference by a constant and determining if the actual value is greater or less than the estimate by more than the allowed amount. As with the generation of the Normalcy (n) measure, the calculation is relatively simple and well known in the art, and therefore any routine software implementation for a standard single linear regression will suffice. Once the standard single linear regression for the entity is calculated, the weight for the change measure for this criteria is looked up in the weight table 46 and then applied to (d) to find the weighted change score (sd). This (sd) is stored in the statistical value memory 28. (Fig. 6 gives a detailed block flow diagram of the process by which the weighted change (sd) score is generated.)

Returning to Fig. 2, the entity-criterion score generator 30 generates a single score for each entity-criterion. The entity-criterion score generator 30 retrieves the (sv), (sn) and (sd) scores for each entity-criterion and sums them to form a single entity-criterion score. The weight for each entity-criterion score for each entity is looked up in the weight table 46 and is then applied to the entity-criterion score to find the weighted entity-criterion score which is stored in the entity-criterion memory 32.

The entity score generator 34 takes each of the weighted entity-criterion scores from the entity-criterion

memory 32 and sums all the weighted entity-criterion scores to form a single entity score. This score is then stored in the entity score database 36.

The risk analysis processor 38 takes the single entity score for all entities processed and compares them to a threshold value derived from the threshold table 48. The entities are then given a label corresponding to their amount of risk in relation to the threshold value.

As shown in Fig. 3, the entities are displayed by the risklist display 40, which uses text, tables, graphs, and graphical indicators to identify the levels of risk for each entity with respect to their engagement in fraudulent activities. In the preferred embodiment, entities are identified with a "red ball" to indicate high risk, a "yellow ball" to indicate medium risk, and a "green ball" to indicate low risk. This risklist of entities may be provided to an appropriate printer/plotter device 54 or display 52 to show the degree of risk for each entity.

One object of the present invention is to provide information to a human analyst in order to assist the analyst in making decisions and taking actions on appropriate risk entity targets. The risk action generator 42 offers a list of actions that can be taken against an entity. In the preferred embodiment, these actions can include a notice to the provider of performance aberration, or a request for an explanation of the behavior in a particular area. Actions are carefully defined to be implementable in enterprise operations and traceable within the system. A generic list of actions would include: monitor, notify, warn, threaten, and cancel.

The analyst user of the fraud pre-processor detection system scans the results of the analysis from the risklist display 40 and selects entities that appear to exhibit the highest risk of fraud. The analyst decides whether an action should be taken against an entity. The purpose of taking action is twofold: (1) to induce a change in the entity's behavior that is observable by the system and which will indicate his or her prior fraudulent actions (by highlighting those behavior changes in subsequent analysis); and (2) to modify charges in such a way that savings produced by the fraud pre-processor detection system can be detected and commented upon. The system can also provide support to the conduct of investigations as a result of the observations and conclusions that are developed.

If an action is selected by the analyst user, the actions taken against a particular entity are recorded and monitored by the risk behavior monitor 44. The risk behavior monitor 44 stores the record of the actions, annotations that may be recorded by the analyst, the identity of the analyst, and other information important to maintaining knowledge of the actions related to each entity in the entity status database 47.

The summary and savings generator 45 provides automatic review of the results of the system on a monthly, or other selected period, basis. The summary and savings generator 45 summarizes the number of entities reviewed in each category along with the total amount of dollars involved. Entities identified in various levels of risk are summarized, along with the number and type of actions taken. The system tracks actions and measures taken prior

to action having been taken, and subsequent performance for entities after actions have been taken. The summary and savings generator 45 provides a unique apparatus and process for determining the two most critical factors in anti-fraud management: (1) specific aberrant behavior detection; and (2) cost reduction/savings.

The summary and savings generator 45 is connected to a group summary processor 49 and an individual summary processor 51. The group summary processor 49 performs a monthly summary of the statistics relating to the group under review, such as the size of the group, the numer of high risk entities, amount charged by the group, amount charged by the high risk entities in dollars and percent of the total charges, total actions taken by type (e.g., level 1, 2, 3, etc.), the total amount claimed for savings, and so forth.

The individual summary processor 51 works only on those entities that have been identified with an action flag in the entity status database 47 through the risk action monitor 44 in order to perform specific aberrant behavior detection. When the user analyst initiates one of the actions, the level of the action, the date taken, the date to initiate subsequent review and the evaluation period to review, are all recorded in the entity status database 47.

The individual summary processor 51 operates on a general assumption that most individuals who are knowingly doing something wrong, will, when attention is called to this fact by an external source, modify their behavior to reduce their risk of discovery and/or punitive action.

Consequently, the apparatus and process used in the invention focuses on behavior modification and the subsequent detection of the changes making the assumption that there is a high degree of correlation between the observed changes and the areas of the entity's conduct that were either fraudulent or, at least, suspect.

Individual entities are screened and scored for their "risk" regarding fraud potentials during an initial pass, or first review pass, through the system. The user analyst performs a review of the output scores in the measured criteria generated by the risklist display 40 and determines that some action generated by the risk action generator 42 should be taken against an entity. He uses the risk action monitor 44 to record the action taken. The entity's record is attached with the appropriate action flag, along with the dates of action and the dates for subsequent review and added to the entity status database 47. The normal period of delay from action to subsequent review is 90 days.

The actions generated by the risk action generator 42 are generally a range of options such as: (1) monitor (i.e., place behavior under continuing review without notifying the individual); (2) notify (i.e., communicate with the individual that general behavior patterns have been detected that appear to be in question - but provide only limited details); (3) warn (i.e., communicate to the individual that specific patterns of questionable behavior have been determined, indicating the nature of the behavior and suggesting self review); (4) threaten investigation (i.e., communicate that unacceptable behavior has been detected and that unless corrected

immediately detailed investigation will result); (5) initiate investigation (i.e., communicate that unacceptable behavior of such dimension has been detected that an investigation is being opened and the individual can expect additional action within a specified period of time); and (6) place on probation (i.e., unless detected unacceptable behavior is changed within a specified time, the individual will be dropped as a provider of services/products).

While the exact wording or communications will be unique to each client application, this generic range of actions will be used and the risk action monitor 44 will record each action implemented in the entity status database 47. Since the actions are coded and graded for increasing degrees of severity, the subsequent monitoring of individual performance will take into account the severity of the action when scoring response behavior. For example, a large response (i.e., change in behavior) from a relatively low level action (e.g., #2 - Notify) would be scored higher than a low response for the same action, or for a higher severity action.

The individual summary processor 51 determines which entities to review from the entity status database 47. It also determines from the date of the action and the delay period when a second pass review of the entities should be performed. Once an entity is selected for a second pass review, the individual summary processor 51 initiates the system to analyze client transaction data from the client data source 12 from the date the action was taken to the date when the second pass review was ordered, for that particular entity. The individual summary processor 51 automatically retrieves the prior data on the

entity saved from the first pass review along with the results of the second pass review. The separate passes are extracted from the 36 entity score database as distinct performance periods on either side of the action-taken date. The data for the initial performance period and the subsequent performance period are compared for differences. Downward changes for charges are recorded as potential savings resulting from the action taken. The changes in specific criteria of measurement (e.g., in the preferred embodiment, the number of patients seen; the number of claims submitted; particular procedures rendered) are detected and used as an indicator of potential areas of fraud in the initial (i.e., first pass review) behavior period. Since the severity of the action taken will reflect on the degree of change to be expected, a multiplier from the 46 weight table is applied to amplify the differences observed according to the action taken. Also, since the prior performance of the entity with respect to his or her change trend will bear on the validity of the difference findings, a trend factor from the trend factor table 53 is applied. If the entity's prior performance is highly variable, the validity of the difference findings is reduced. Conversely, if the prior behavior trend is found to be very stable, significant differences will be amplified.

The results of the behavior change analysis performed by the individual summary processor 51 are presented to the analyst. Where the difference results indicate a high degree of correlation with the risk criteria identified in the first review pass, the analyst can assume a causal relationship with the action and that the areas that have changed are suggestive of aberrant

behavior prior to the action. The analyst can then use the behavior changes as a means of reviewing highly specific areas of the entity's prior performance, as well as modify the weight and threshold values found in the 46 weight table and 48 threshold table to better refine the detection process for subsequent reviews. The technique embodied in the invention provides several unique and previously unavailable capabilities to the process of fraud detection.

As shown in Figs. 4, 4A, 5 and 6, details of the block flow diagrams showing the steps for generating the statistical measures for Value (v), Normalcy (n) and Change (d) are shown. As the means for generating these statistical measures are well-known in the art, they will not be described in detail herein. It should be noted, however, that although the means for generating these statistical measures are well-known, the application of these statistical measures to transactional data in order to analyze such data for purposes of fraud detection is considered a novel feature of the invention.

As shown in Fig. 7, which details a block flow diagram showing the steps for performing the pre-processor method, the initial stages of the process serve to filter and highlight suspect individuals, as described previously. Once suspect individuals are detected, prioritized and highlighted by the system an analyst will review the results and actions generated by the risklist display 40 and risk action generator 42, respectively, to determine what action to take. This automated filtering and prioritization process is critical because reviewing all of the behavior of all entities in sufficient detail to precisely identify fraud is impractical. The amount of

data normally collected in the transaction-billing process is far to great to permit manual, or human review. And the potential variations for fraud in such industries as health care are far too many, and much too flexible, to build a completely automated pattern detection system that would be reliable for any protracted period. Consequently, the present invention reduces the focus of the human analyst to the most likely payoff.

The combined statistical approach of the pre-processor apparatus 110 highlights the most critical areas of behavior in performing the filtering, that is: value, normalcy, and change. These measures, when applied to carefully selected behavior criteria (i.e., selected for their potential to exhibit profit-taking) produce a high degree of validity in showing the analyst where to look. The output of the first pass review is normally sufficient for the analyst to decide to take action.

This element of the present invention is one of the unique components of the process - the concept of introducing process control theory to the detection of fraud in large transaction data histories. There has not, heretofore, been available a methodology that was implementable in an automated process that could record behavior changes in specific entities and determine the savings resulting from the counter-fraud program. As a result of the first pass review, organizations currently suffering from high losses due to fraud have a means of not only identifying candidate targets, but they can modify the behavior of those targets in order to obtain savings and further focus on specific fraudulent behavior. The results of the process can offer the analyst with the option of

declaring savings or using the highlighted areas of change as a means of initiating deeper investigation into very specific areas of the entity's performance.

While the initial use of the process is focused on fraud in the health care industry, it is apparent that the technique described here can also be used to detect fraud and identify resulting savings in a variety of environments (e.g., banking, mortgage, worker's compensation, etc.).

I CLAIM:

1. A fraud detection system for targeting a potentially fraudulent entity within a particular industry, comprising:

statistical analysis engine for generating statistical values for said entity according to at least one of a plurality of analysis criteria;

entity-criterion score generator for weighting and summing each of said statistical values thereby forming an entity-criterion score for each of said analysis criteria;

entity score generator for weighting and summing each said entity-criterion scores to form a first entity score;

computer memory for storing said statistical values, said entity-criterion scores, and said first entity score; and

risk analysis processor for comparing said first entity score to a predetermined threshold, whereby the result of said comparison indicates whether said entity is engaging in fraudulent activity.

2. The system according to claim 1, further comprising controller means for coordinating, controlling, monitoring and executing commands necessary for operating said fraud detection system.

3. The system according to claim 1, further comprising a database for storing said first entity score.

4. The system according to claim 1, wherein said statistical values include a value statistic (v), an normalcy statistic (n), and a change statistic (d) for each of said criteria.

5. The system according to claim 1, further comprising client data converter for converting client records containing data elements useful in measuring the performance of said entity to a format compatible with said fraud detection system.

6. The system according to claim 1, further comprising a three dimensional array for organizing and storing said entity and said criteria according to predefined time units, said array comprising an x, y and z axis, said y-axis containing said entity, said x-axis containing said criteria, and said z-axis containing said time units.

7. The system according to claim 6, wherein said array creates entity-criterion buckets wherein said statistical values for each criterion can be generated for said entity within said time period.

8. The system according to claim 1, further comprising a risklist display for displaying in graphical form whether said entity is engaging in fraudulent behavior.

9. The system according to claim 1, further comprising:

risk action generator for displaying possible actions to initiate against said entity, and selecting one of said actions to modify said entity's behavior;

risk behavior monitor for storing a record of said selected action taken against said entity; and

summary and savings generator for identifying the amount of money gained or lost by said action.

10. The system according to claim 9, wherein said summary and savings means include:

a second entity score presenting said entity's behavior after said action has occurred;

individual summary processor for comparing said first entity score with said second entity score, measuring the degree of change therefrom, weighting the degree of change using a severity factor and a trend factor, and displaying a result indicating whether said action decreased the fraudulent activity of said entity;

group summary processor for generating group statistics for a plurality of entities.

11. The system according to claim 10, wherein said individual summary processor includes adjuster for adjusting said weighting values for said criteria scoring means and said entity scoring means, and said threshold

values for said risk analysis means, in accordance with said group statistics from said summary and savings means.

12. A fraud detection system for targeting potentially fraudulent entities within a particular industry, comprising:

client data source for in putting client records containing data elements useful in measuring the performance of an entity by evaluating a plurality of criteria over a set of time period;

there dimensional array for organizing and storing said entity, said criteria, and said time period, comprising an x, y and z axis, said y-axis containing said entity, said x-axis containing said criteria, and said z-axis containing at least one of said time units;

statistical analysis engine for generating a value statistic (v), a normalcy statistic (n), and a change statistic (d) for each of said criteria;

entity-criterion score generator for weighting and summing each of said statistical values thereby forming an entity-criterion score for each of said criteria;

entity score generator for weighting and summing each of said entity-criterion scores to form an entity score for said entity;

risk analysis processor for comparing said entity score with pre-defined threshold values; and

risklist display for displaying in a graphical manner whether said entity is engaging in fraudulent behavior.

13. A computer program for targeting potentially fraudulent entities within a particular industry, said computer program comprising the steps of:

inputting client records containing data elements useful in measuring the performance of an entity by evaluating a plurality of criteria over a set time period;

organizing and storing said entity, said criteria, and said time period, into a three dimensional array comprising an x, y and z axis, said y-axis containing said entity, said x-axis containing said criteria, and said z-axis containing at least one of said time units;

generating a value statistic (v), a normalcy statistic (n), and a change statistic (d) for each of said criteria;

weighting and summing each of said statistical values thereby forming an entity-criterion score for each of said criteria;

weighting and summing each of said entity-criterion scores to form an entity score for said entity;

comparing said entity score with pre-defined threshold values; and

displaying graphically whether said entity is engaging in fraudulent behavior.

14. The computer program according to claim 13, said computer program further comprising the step of encoding said computer program on a computer-readable medium.

15. A method of targeting potentially fraudulent entities within a particular industry, comprising the steps of:

generating statistical values for entity according to a plurality of analysis criteria;

weighting and summing each of said statistical values thereby forming an entity-criterion score for each of said analysis criteria;

weighing and summing each of said entity-criterion scores to form an entity score; and

comparing said entity score to a predetermined threshold.

16. The computer program according to claim 15, said computer program further comprising the step of encoding said computer program on a computer-readable medium.

17. A fraud pre-processor system for filtering and highlighting those individuals most likely to be engaging in fraudulent behavior, comprising:

a controller;

a computer program executed by said controller comprising:

client data converter for converting client data to format compatible with said fraud pre-processor system;

three dimensional array for storing a plurality of entities, analysis criteria, and time units such that for each of said entities statistical values for each of said criteria within each of said time units can be generated, weighted and summed to form an entity-criterion score;

entity-criterion score generator for generating said entity-criterion score;

entity score generator for weighting and summing each entity-criterion score for each entity to form an entity score;

risk analysis processor for comparing said entity score for each of said entities with a threshold value;

risklist display for displaying the results from said comparison in a manner identifying those entities most likely engaging in fraudulent behavior; and

a database for permanently storing, and a computer memory for temporarily storing, said analysis criteria, said statistical values, said entity-criterion score, said entity-score, said threshold values, and said results.

18. A fraud detection system, comprising statistical analysis engine for generating statistics for a plurality of entities according to a plurality of analysis criteria and time periods; and

pre-processor for evaluating said statistics to determine if said entities are engaging in fraudulent behavior.

19. The system according to claim 1, wherein said statistical analysis engine further comprises:

value processor for generating said value statistic (v) which determines the damage potential for said entity;

normalcy processor for generating said normalcy statistic (n) for measuring normal behavior for said entities; and change processor for generating said change statistic (d).

20. A fraud detection system, comprising:

risk analysis processor for comparing an entity score for an entity to a predetermined threshold;

risk action generator for displaying and selecting action to be taken said entity;

risk action method for storing a record of said actions; summary and savings generator for identifying the amount of money gained or lost by said action; and

adjustor for adjusting weighting values used to generate said entity score, and said threshold values for said risk analysis means, in accordance with said statistics from said summary and savings generator, in order to better estimate if said entity is engaging in fraudulent behavior.

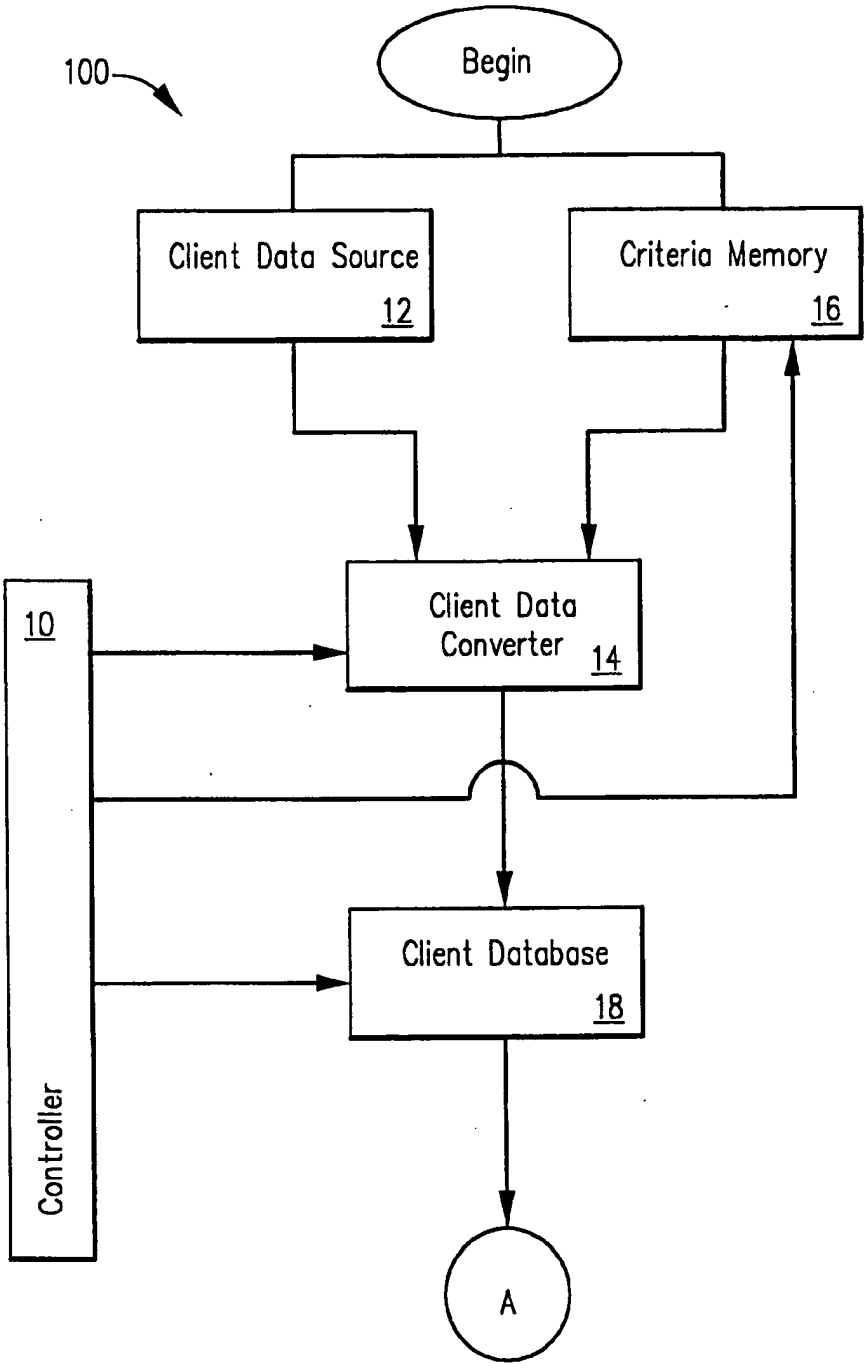

100
Begin
Client Data Source
12
Criteria Memory
16
Client Data
Converter
14
10
Client Database
18
Controller
A

FIG. 1

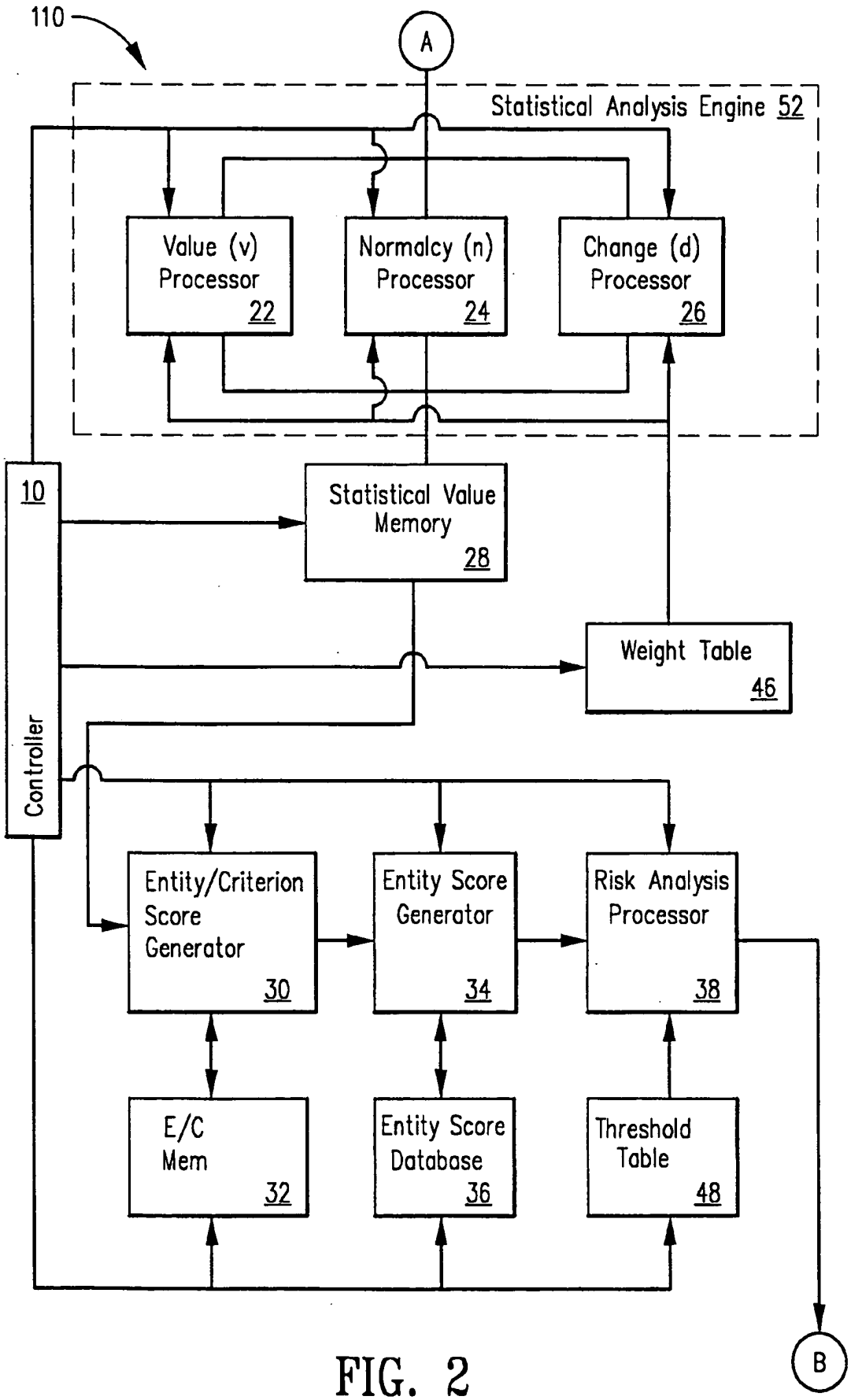

110
A
Statistical Analysis Engine 52
Value (v) Processor 22
Normalcy (n) Processor 24
Change (d) Processor 26
10
Statistical Value Memory 28
Weight Table 46
Controller
Entity/Criterion Score Generator 30
Entity Score Generator 34
Risk Analysis Processor 38
E/C Mem 32
Entity Score Database 36
Threshold Table 48
B

FIG. 2

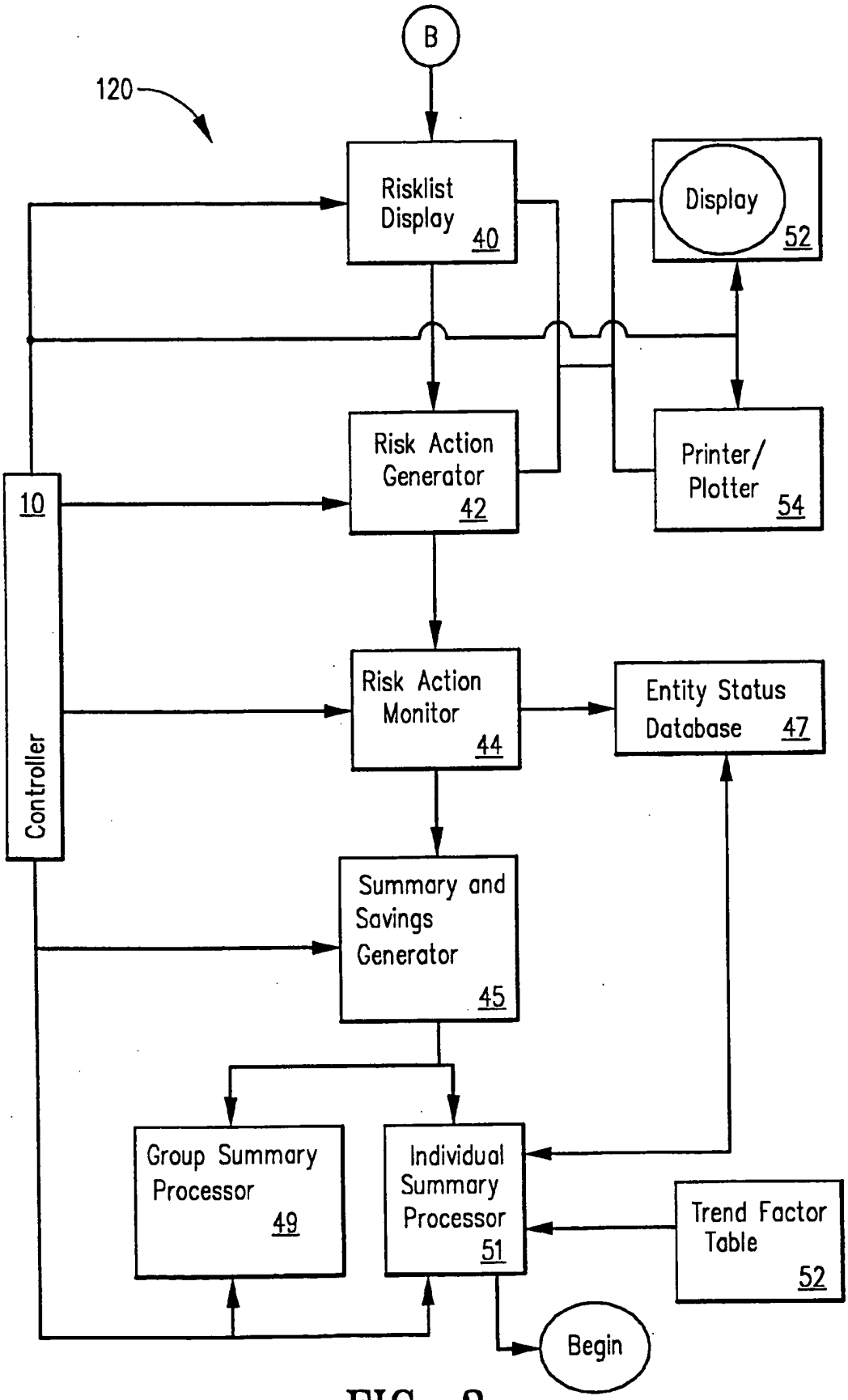
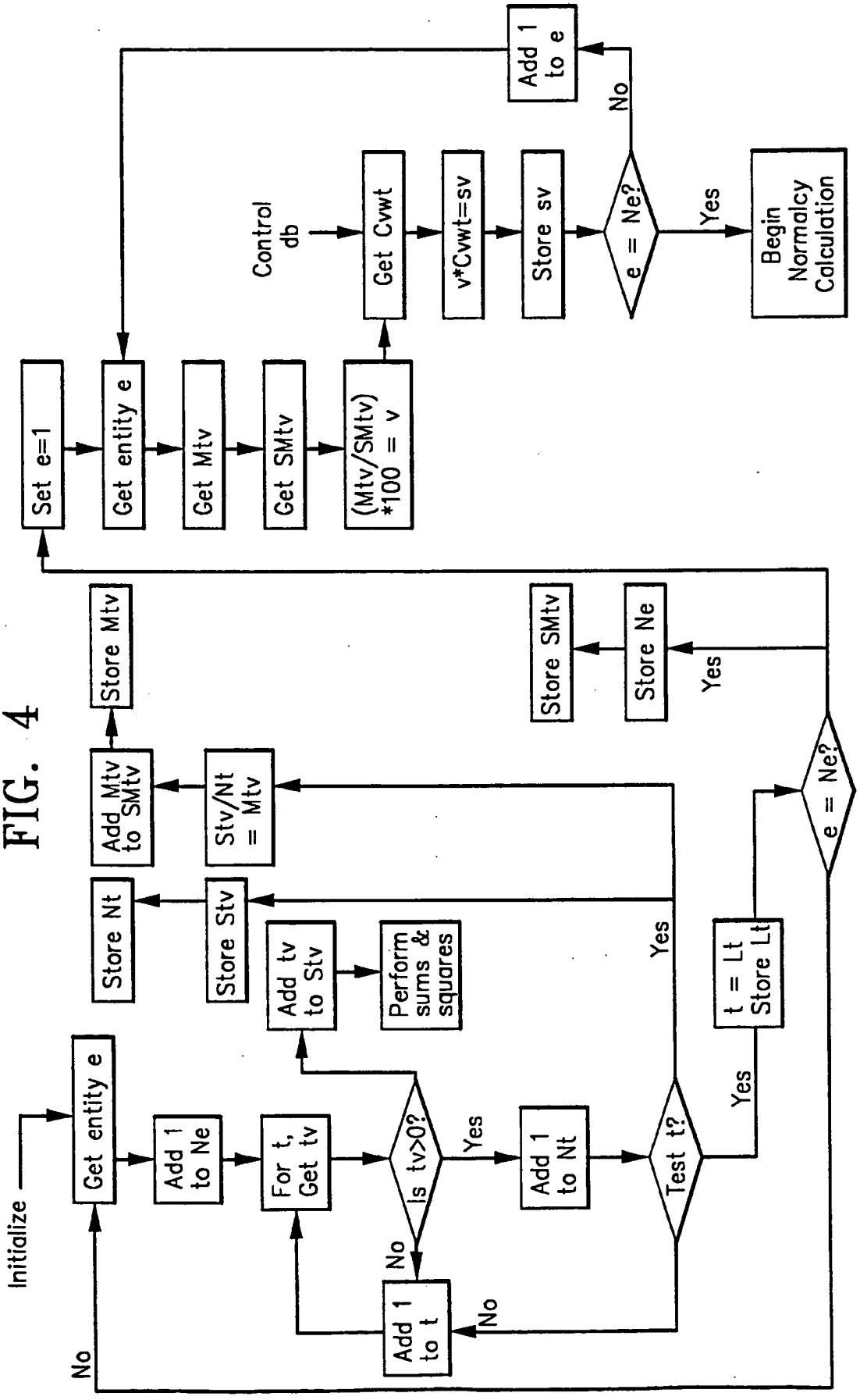

B
120
Risklist Display 40
Display 52
Risk Action Generator 42
Printer/ Plotter 54
10
Controller
Risk Action Monitor 44
Entity Status Database 47
Summary and Savings Generator 45
Group Summary Processor 49
Individual Summary Processor 51
Trend Factor Table 52
Begin

FIG. 3

FIG. 4


Initialize
Get entity e
Add 1 to Ne
For t, Get tv
Is tv>0?
Yes
No
Add 1 to Nt
Add 1 to t
Test t?
Yes
No
Yes
t = Lt Store Lt
e = Ne?
Yes
No
Add tv to Stv
Perform sums & squares
Store Nt
Store Stv
Add Mtv to SMtv
Stv/Nt = Mtv
Store Mtv
Store SMtv
Store Ne
Yes
Set e=1
Get entity e
Get Mtv
Get SMtv
(Mtv/SMtv) *100 = v
Control db
Get Cvwt
v*Cvwt=sv
Store sv
e = Ne?
Yes
No
Add 1 to e
Begin Normalcy Calculation

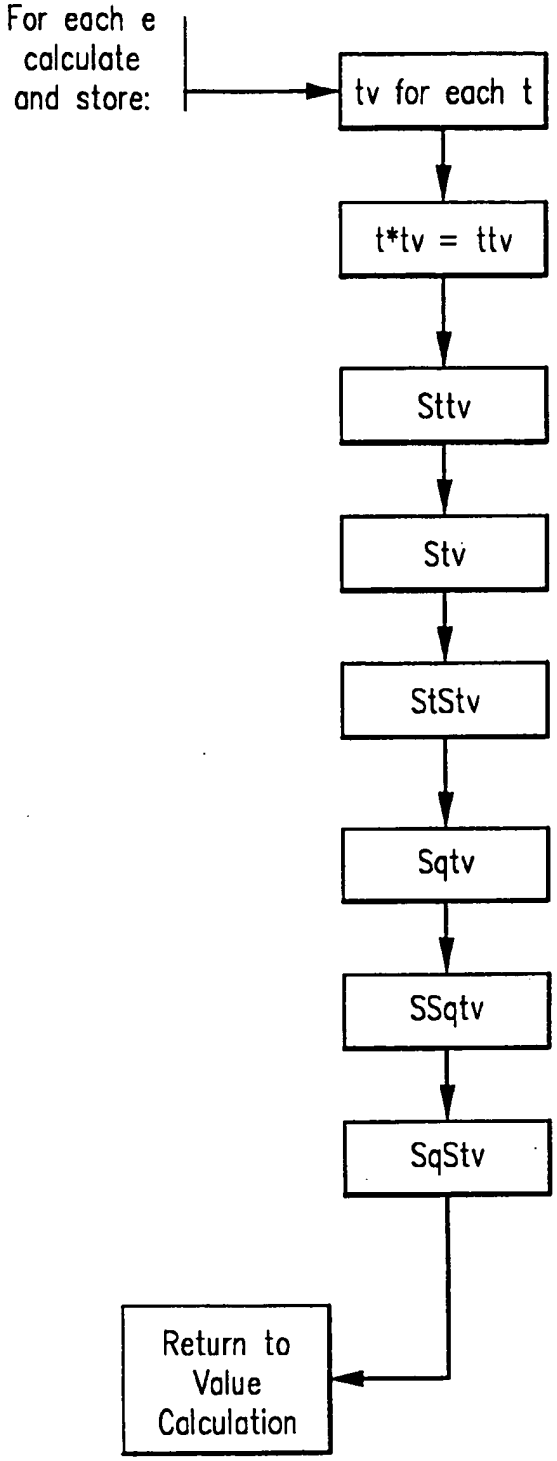

For each e
calculate
and store:
tv for each t
t*tv = ttv
Sttv
Stv
StStv
Sqtv
SSqtv
SqStv
Return to
Value
Calculation

FIG. 4A

FIG. 5

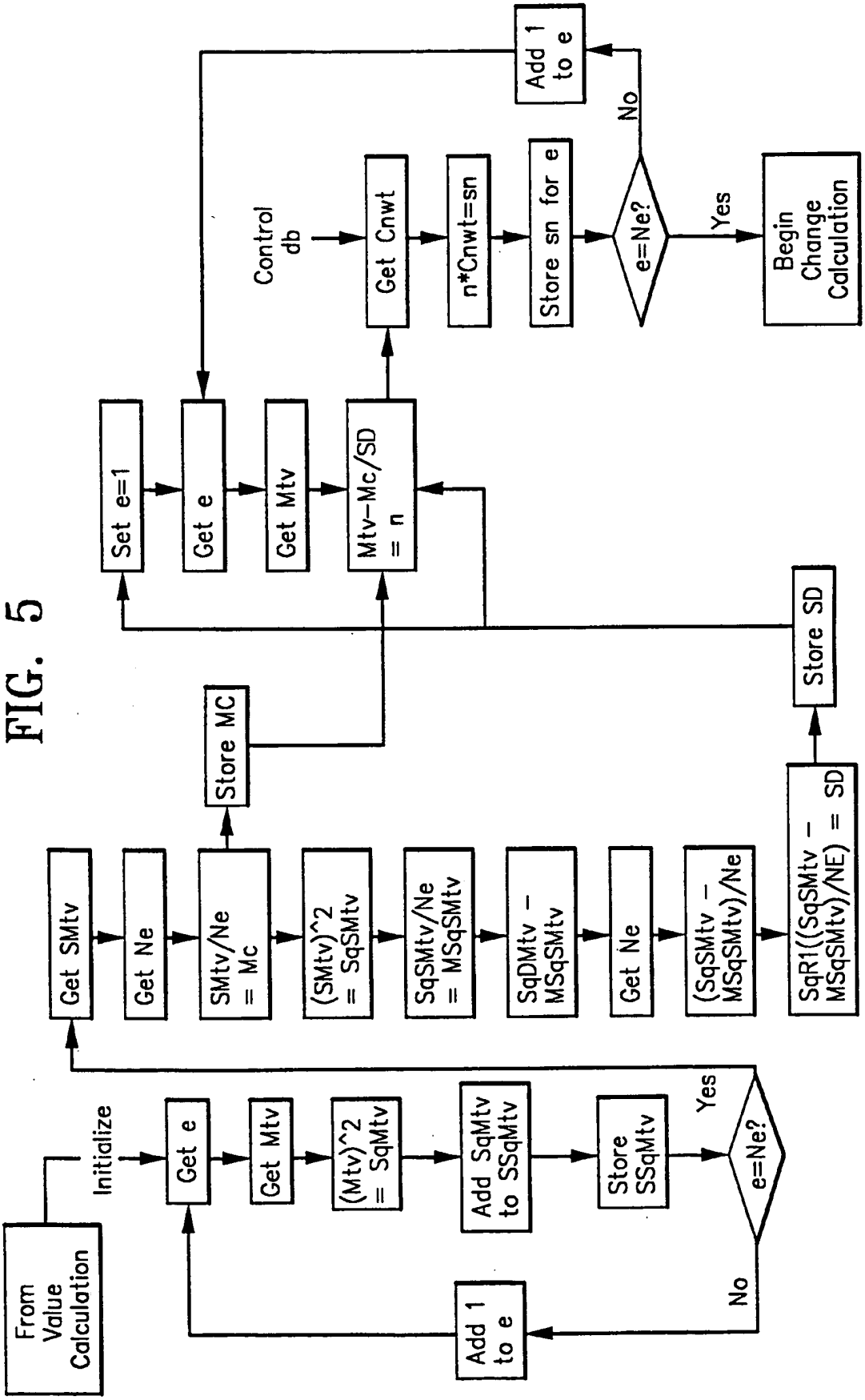
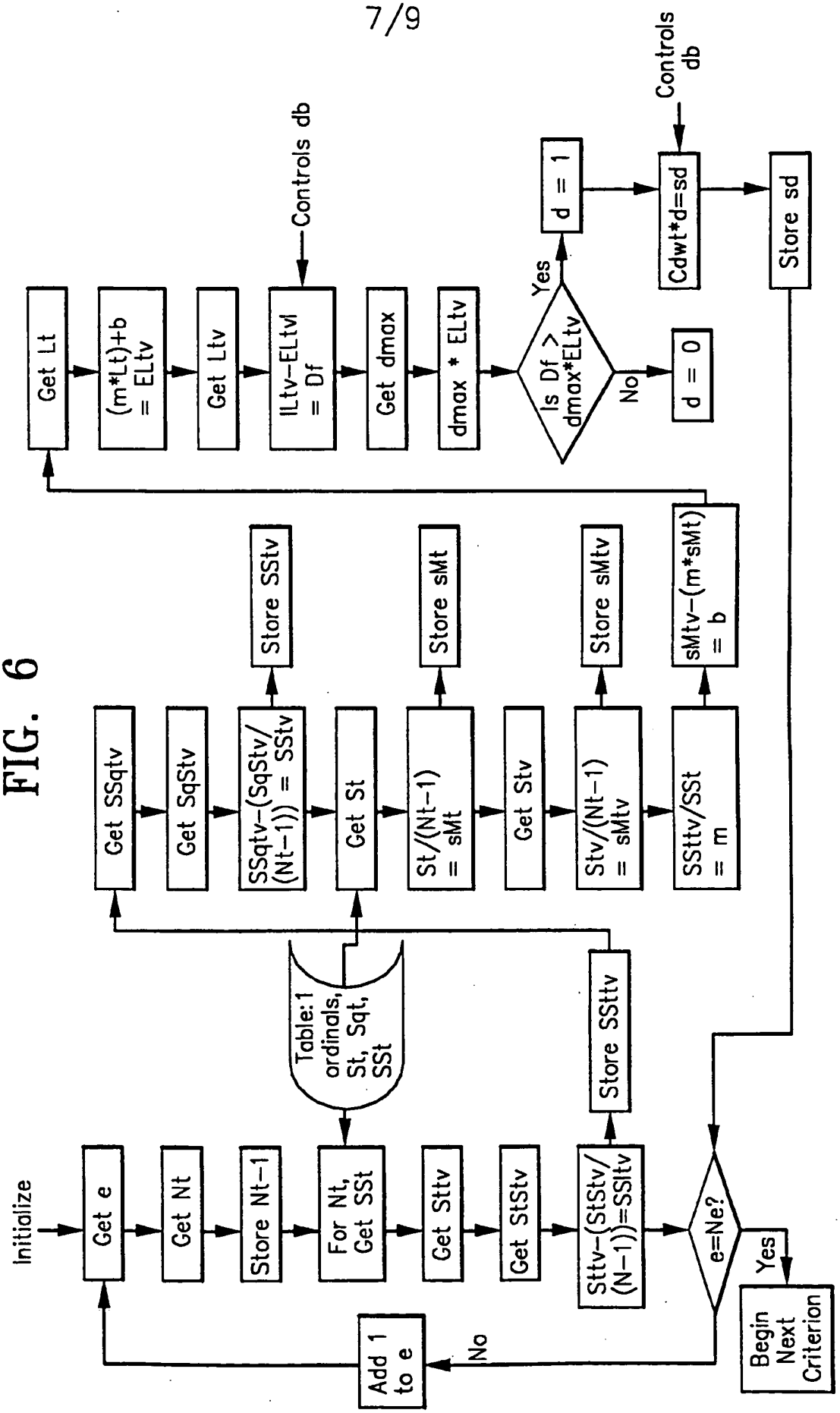

From Value Calculation
Initialize
Get e
Get Mtv
(Mtv)^2 = SqMtv
Add SqMtv to SSqMtv
Store SSqMtv
e=Ne?
Yes
No
Add 1 to e
Get SMtv
Get Ne
SMtv/Ne = Mc
(SMtv)^2 = SqSMtv
SqSMtv/Ne = MSqSMtv
SqDMtv - MSqSMtv
Get Ne
(SqSMtv - MSqSMtv)/Ne
SqR1((SqSMtv - MSqSMtv)/NE) = SD
Store SD
Store MC
Set e=1
Get e
Get Mtv
Mtv-Mc/SD = n
Control db
Get Cnwt
n*Cnwt=sn
Store sn for e
e=Ne?
Yes
No
Add 1 to e
Begin Change Calculation

FIG. 6


Initialize
Get e
Get Nt
Store Nt-1
For Nt, Get SSt
Get Sttv
Get StStv
Sttv-(StStv/(N-1))=SSltv
Store SSttv
e=Ne?
Yes
No
Add 1 to e
Begin Next Criterion
Table:1 ordinals, St, Sqt, SSt
Get SSqtv
Get SqStv
SSqtv-(SqStv/(Nt-1)) = SStv
Store SStv
Get St
St/(Nt-1) = sMt
Store sMt
Get Stv
Stv/(Nt-1) = sMtv
Store sMtv
SSttv/SSt = m
sMtv-(m*sMt) = b
Get Lt
(m*Lt)+b = ELtv
Get Ltv
Controls db
lLtv-ELtvl = Df
Get dmax
dmax * ELtv
Is Df > dmax*ELtv
Yes
No
d = 1
d = 0
Controls db
Cdwt*d=sd
Store sd

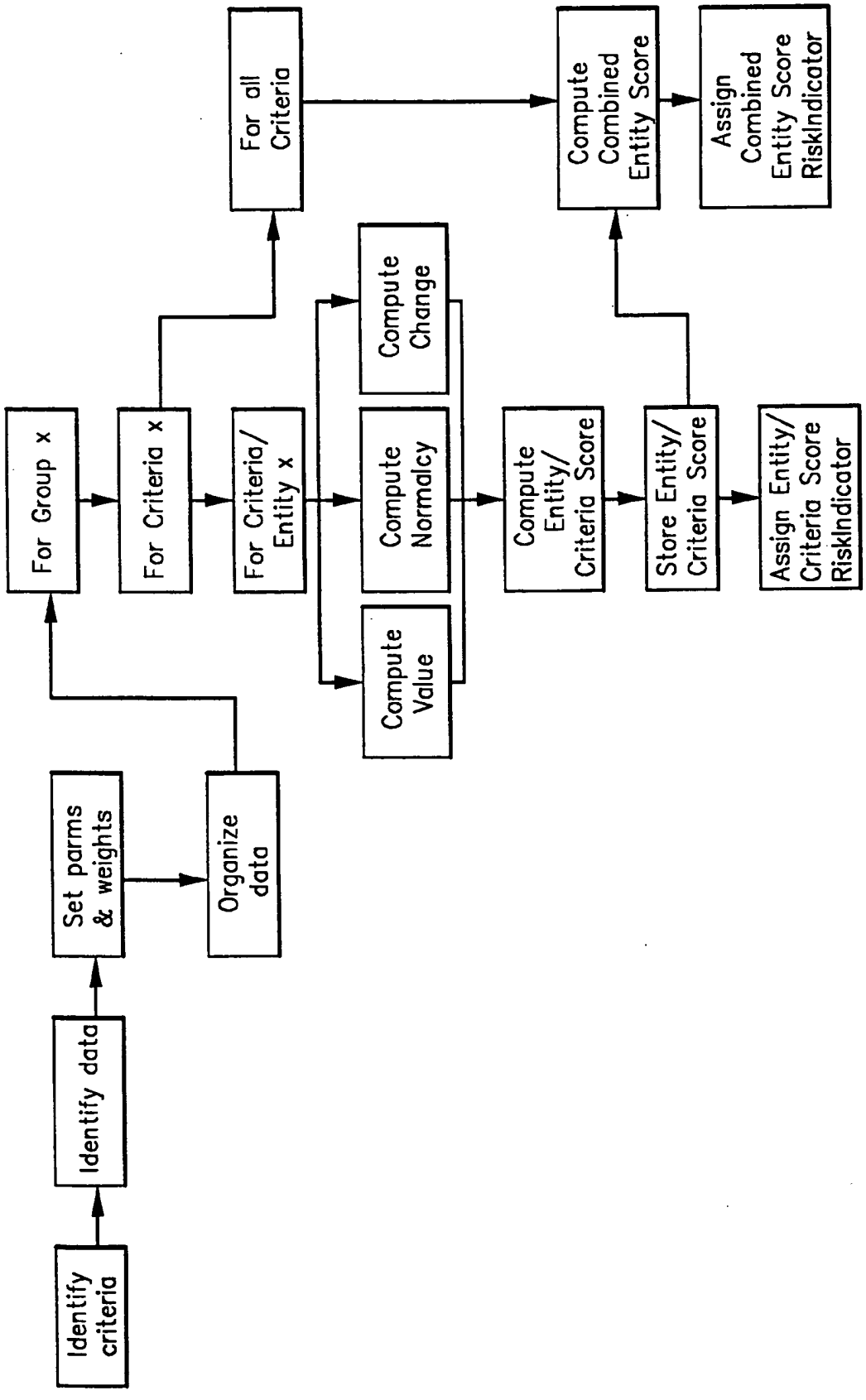

Identify criteria
Identify data
Set parms & weights
Organize data
For Group x
For Criteria x
For Criteria/ Entity x
Compute Change
Compute Normalcy
Compute Value
Compute Entity/ Criteria Score
Store Entity/ Criteria Score
Assign Entity/ Criteria Score RiskIndicator
For all Criteria
Compute Combined Entity Score
Assign Combined Entity Score RiskIndicator

FIG. 7

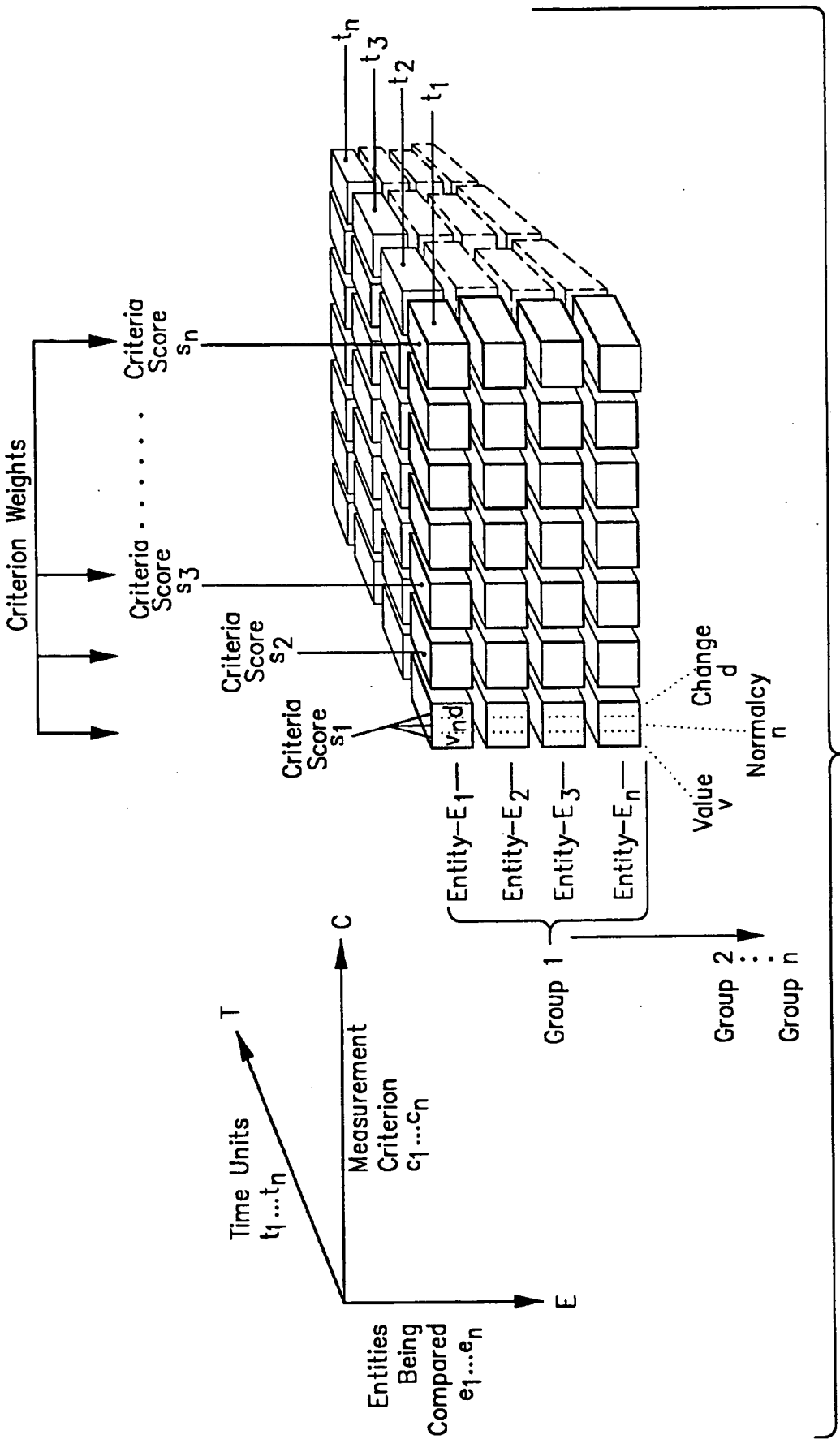

Criterion Weights
Criteria Score s_n
Criteria Score s_3
Criteria Score s_2
Criteria Score s_1
t_n
t_3
t_2
t_1
v_{nd}
Change d
Normalcy n
Value v
Entity-E_1
Entity-E_2
Entity-E_3
Entity-E_n
Group 1
Group 2
Group n
T
C
E
Time Units $t_1 \ldots t_n$
Measurement Criterion $c_1 \ldots c_n$
Entities Being Compared $e_1 \ldots e_n$

FIG. 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/10352

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :G06F 15/00, 17/60; G06G 7/52
US CL :364/401M, 401R, 406, 408
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 364/401M, 401R, 406, 408

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y, E	US, A, 5,557,514 (SEARE ET AL) 17 September 1996, cols. 4, 19, 21	1-20
Y	US, A, 5,253,164 (HOLLOWAY ET AL) 12 October 1993, col. 3	1-20
Y	US, A, 5,359,509 (LITTLE ET AL) 25 October 1994, cols. 4,5,8,9,13,14	1-20

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
22 OCTOBER 1996

Date of mailing of the international search report
27 NOV 1996

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
GAIL O. HAYES
Telephone No. (703) 305-9711

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/10352

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS, DIALOG, STN
Search Terms: track? or monitor? or detect?; fraud? or aberration? or deviat?; historic? or past or previous; analys? or credit? or check?

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :		(11) International Publication Number:
G06F 15/00, 17/60, G06G 7/52	A1	WO 97/00483
		(43) International Publication Date: 3 January 1997 (03.01.97)

(21) International Application Number: PCT/US96/10352

(22) International Filing Date: 14 June 1996 (14.06.96)

(30) Priority Data:
08/490,984 15 June 1995 (15.06.95) US

(71) Applicant: FRAUDetect, L.L.C. [US/US]; Suite 312, 10400 Eaton Place, Fairfax, VA 22030 (US).

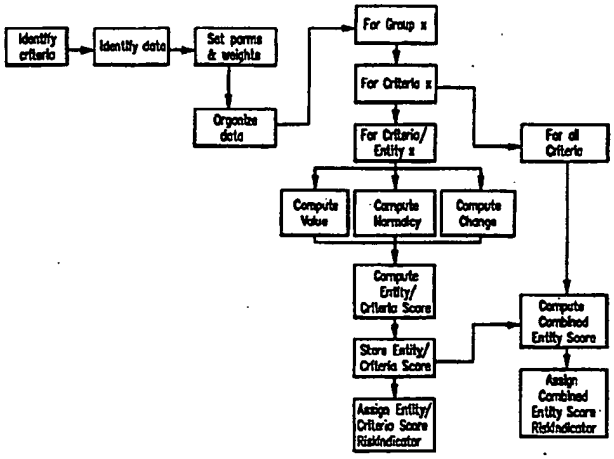
(72) Inventor: COFOD, Robert, K.; 282 Prince's Lane, Harwood, MD 20776 (US).

(74) Agent: GROSSMAN, Jon, D.; 2101 L Street, N.W., Washington, DC 20037 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: PROCESS AND APPARATUS FOR DETECTING FRAUD



(57) Abstract

A system and method for processing transaction data in filtering and targeting potentially fraudulent entities within a particular industry relies upon client data input apparatus for inputting client/provider/entity data records containing data elements for measuring an entity's performance according to pre-selected criteria, where entity date, criteria data and time data are organized and stored in three-dimensional array. A pre-processor (110) generates a score for each entity-criterion (1...n) by generating a value statistic (v), a normalcy statistic (n), and a change statistic (d), weighting each statistical value, and summing the values. An entity scoring generator (30) generates a single entity score for each entity by weighting each entity-criterion score and summing the entity-criterion scores. The entity scores are compared to pre-defined threshold values for detecting those entities most likely to be engaging in fraudulent activity and a graphical display entities according to level of risk.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

PROCESS AND APPARATUS FOR DETECTING FRAUD

FIELD OF THE INVENTION

The invention relates to a process and apparatus for analyzing client data against pre-selected criteria over a pre-determined period of time in order to identify those individuals or entities engaging in activity indicative of potential fraudulent behavior within a particular industry.

BACKGROUND OF THE ART

Many industries, such as medical and banking, suffer enormous losses due to the theft of money or services through fraudulent means, such as submitting insurance claims for unseen patients or forging checks. In the medical industry, for example, it is estimated that approximately $100 billion of U.S. annual health care expenditures is stolen through fraud or abuse. Even more alarming is that some experts believe this figure only represents the portion of which we are aware. The actual figure may be significantly higher, perhaps as much as twice the amount.

The problems surrounding fraud detection are many. Since most companies maintain relatively few investigators, limited resources are available to perform the analysis and decision-making necessary to accurately attack the problem and ensure measurable results. In addition, the possibility of wrongly accusing a client of fraudulent behavior adds an extra burden of caution and proof. Confidentiality, government involvement, market

competition, and judicial processes add other limits to the options for attacking the problem.

Other problems in fraud detection include highly varying types of targets, possible scenarios, and innovative fraud techniques requiring heavy reliance on skilled human analysts and decision makers to detect, recognize, define and initiate actions. In addition, each industry, such as the aforementioned medical and banking industries, generates a massive and dynamic quantity of information from which relevant "target" information must be extracted. Furthermore, there is a time sensitivity factor in processing and acting upon newly detected targets or instances of fraud.

Due to the above-cited difficulties in detecting those persons engaging in fraudulent activities, billions of dollars worth of stolen money or services have largely been ignored or passed on as a cost of operation.

Most cases of fraud are discovered as a result of tip-offs or indications from industry program integrity reviews. These discoveries have often been cited as being only the tip of the iceberg, yet most investigative staffs are already inundated with the cases on hand.

The problem is not unlike another well-known problem in military intelligence and targeting circles. It is known as a "target rich environment," where the number of possible targets far exceeds the resources available for attack. The military problem, like the one faced in combating health care and banking fraud and abuse, is fundamentally one of resource management. The problem

requires a method or system by which an industry may increase ensurance that the resources available to it are being spent on targets with the highest potential payback. Consequently, there is a need for processing, ordering and analyzing the volumes of transactional data, such as billing information and insurance claims, associated with a particular entity, such as a physician working within the health care industry, and extrapolating from such information whether the client is engaging in fraudulent behavior. This need for specifically targeting perpetrators of fraud or abuse becomes more readily apparent in light of the fact that there may be tens or hundreds of thousands of clients per industry. Accordingly, the present invention answers that specific need.

SUMMARY OF THE INVENTION

It is therefore apparent from the above that there exists a significant need for a system and process which effectively uses transactional data in a manner that identifies potentially fraudulent behavior and areas of greatest need for investigation. It is therefore one of the objectives of this invention to provide a fraud pre-processor detection system. Fraud can be detected for many industries dealing with a large number of clients, such as the aforementioned medical and banking industries.

Another object of this invention is to provide a process for analyzing large volumes of transaction data for the purpose of detecting indications of fraud.

Another object of this invention is to provide a system and process which automatically prioritizes those

clients or entities which are most likely to be engaging in fraudulent behavior.

Another object of this invention is to provide a system and process for establishing thresholds indicating fraud.

An additional object of this invention is to provide a system and process for filtering, processing and displaying data in a manner where it may be used as an analysis tool to support the human decision making process.

Another object of this invention is to provide a system and process for organizing data in a three dimensional array consisting of pre-selected criteria along the x-axis, groups of entities along the y-axis, and time units along the z-axis, thereby allowing entity-criterion analysis, that is analyzing each entity within a group (a single chiropractor out of a group of chiropractors) according to each criterion (e.g., the number of patients seen by the chiropractor, or the amount of claims submitted to the insurance carrier), over a pre-selected period of time units (e.g., days, months or years).

Yet a further object of this invention is to provide a system and process for generating statistical values for each entity-criterion to measure Value (v), or the relative importance of the entity with respect to the entire group, Normalcy (n), or how the entity conforms to the normal behavior of the group for a particular criterion, and Change (d), or the amount of change in activity by the entity over a period of time.

It is also an objective of this invention to weight the statistical values for Value (v), Normalcy (n) and Change (d) according to the relative importance of each value within a criterion with respect to targeting entities engaging in fraudulent activity.

Yet a further object of this invention is to add all the statistical values for each entity-criterion together to create a single entity-criterion score, weight the score according to its fraud targeting potential, sum all the individual entity-criterion scores to form a final entity score, and compare that score to a threshold value to determine the risk of fraudulent behavior.

A further object of this invention is to display all the entities under analysis in a manner that allows easy identification of those entities engaging in behavior most likely indicative of fraud.

Another object of the invention is to provide a list of actions to take against those entities suspected of fraudulent behavior so that a human analyst may select the appropriate action.

It is a further object of this invention to provide a system and process for monitoring those entities suspected of engaging in fraud once an action has been taken against the entity, and determining whether the behavior of the entity changes, and measuring such change, thereby providing some evidence that fraudulent behavior was accurately identified and corrected while also providing a mechanism by which the weighting of the relative importance of various criteria, entity-criterion

scores and entity scores can be modified to more accurately detect the presence of fraudulent behavior.

It is yet a further object of this invention to provide a system and process for generating a change meter which can be translated into dollar amounts to calculate savings to a particular industry.

Briefly described, these and other objects of the invention are accomplished by providing a system and process for inputting client transaction data for a particular industry, organizing the client transactional data into a three dimensional array, generating statistical values for each entity-criterion over time by measuring (v), (n) and (d), weighting each statistical value and summing all values to form a single entity-criterion score, weighting each entity-criterion score and summing all entity-criterion scores to form a single entity score, comparing the entity score to a pre-determined threshold, and displaying the entities so that those engaging in fraudulent behavior are easily identified.

The system and process also include providing a list of actions for the human analyst to take against those entities with scores indicating fraudulent behavior, and monitoring the behavior of those entities to see if they change thereby providing some evidence that the entities were initially engaging in fraudulent behavior, as well as providing information to more precisely adjust the weighting and threshold mechanism used by the system.

With these and other objectives, advantages and features of the invention that may become apparent, the

nature of the invention may be more clearly understood by reference to the following detailed description of the invention, the appended claims, and to the several drawings attached herein.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a hardware block diagram of the client data input apparatus of the present invention;

Fig. 2 is a hardware block diagram of the pre-processor apparatus of the present invention;

Fig. 3 is a hardware block diagram of the monitoring and tracking apparatus of the present invention;

Fig. 4 is a block flow diagram showing modules and steps for generating the Value (v) statistic;

Fig. 4A is a block flow diagram showing modules and steps for performing the sums and squares calculation used in generating the Value (v) statistic;

Fig. 5 is a block flow diagram showing modules and steps for generating the Normalcy (n) statistic;

Fig. 6 is a block flow diagram showing modules and steps for generating the Change (c) statistic;

Fig. 7 is a block flow diagram showing modules and steps for a preferred embodiment of the present invention; and

Fig. 8 is a model showing the theoretical data structure used by the present invention.

DETAILED DESCRIPTION OF THE INVENTION

Referring now to the drawings wherein like elements refer to like reference numerals throughout, Figs. 1, 2 and 3 are hardware block diagrams for a fraud pre-processor detection system according to the present invention. An objective of the present invention is to provide a mechanism for analyzing transaction data in filtering and targeting potentially fraudulent activities of individual entities. Entities may include health care providers, bank account holders, or other individuals to be analyzed for fraud risk "targeting". The term targeting refers to isolating from a universe of entities a sub-set of those entities most likely to be engaging in fraudulent activities.

Specifically, Fig. 1 illustrates the client data input apparatus for inputting and organizing large amounts of client transactional data. Fig. 2 illustrates the pre-processor apparatus for evaluating an entity by generating statistical values for each of a number of pre-selected criteria, weighting and summing each statistical value generated to form a single entity-criterion score, weighting and summing each entity-criterion score to form a single entity score, and comparing this entity score to a predetermined threshold. Fig. 3 illustrates the monitoring and tracking apparatus for displaying a "risk list" of entities with indicators as to their level of risk with respect to engaging in fraudulent activities, generating a list of actions to take against those entities with the highest risk, monitoring

those entities against which actions have been taken in order to determine if behavior has been modified, and providing data for further refining weighting and threshold values to more accurately identify fraud in light of such modified behavior.

A preferred embodiment of the data processed by this invention relates to the medical industry wherein transactional data (e.g., billing records or insurance claims) for a particular entity within a group (e.g., a cardio-thoracic surgeon within a group of surgeons or doctors, or a periodontist within a group of dentists, and so on) is analyzed to determine if patterns of fraudulent behavior can be detected. However, the processes and apparatus described herein can be applied to any other client transactional data for any number of industries.

It has been found in the preferred embodiment, that by examining client transactional data according to certain criteria, patterns can be detected. In the health care administration process, data collection occurs in a variety of ways: claims processing, subscriber enrollment, provider certification, reimbursement-accounting, and investigations. In some cases, the data is collected from external sources (e.g., provider certification). In other cases, the data is derived from internal processes (e.g., reimbursement-accounting). In health care, most of the claims/subscriber/provider related data occurs as a normal course of doing business. Also included as "data" is that information developed as a result of internal or external investigation regarding a particular case.

Administrators of large plans may involve millions of subscribers and hundreds of thousands of providers. Consequently, there is a tremendous amount of data, and this aids the perpetrator of fraud or abuse since, with some care, he or she can lie buried and virtually undetected. Because of the economics of the industry, most data processing systems are usually optimized for the rapid processing of claims, thereby limiting the kind of ad hoc analysis that is helpful in combating fraud and abuse.

As a result of the huge amount of client transactional data available for analysis, it is necessary to store in the criteria memory 16 (Fig. 1) a number of analysis criteria for use in the system as indicators of fraud. It has been found that for any domain environment subject to fraud there are many observable criteria that can be found in transaction data and history records. By understanding the operational characteristics of fraud in a particular domain, it is possible to select specific criteria than can be used to detect the risk of fraud. These criteria are determined manually through the user's knowledge of the fraud environment (e.g., how physicians commit fraud) and how the related behavior is represented in the transaction data and history. The objective of selecting the criteria is to identify independent variables that can reflect behaviors that are focused on maximizing income and minimizing investment.

In general, these criteria can include two types: (1) information which can be summed; or (2) information which can be counted. In the medical industry, examples of these criteria can include the number of claims filed by a provider (counted), the total amount billed (summed), the

number of critical versus non-critical patients (counted), the number of given patients seen by a doctor over the period of a day, week or year (counted), and so forth. These criteria can also include specific patterns indicating fraud such as billing charges for intensive care ambulance services for a non-critical patient. These specific patterns, however, address only a small number of the possible ways fraud or abuse can be committed, and are therefore of limited use. Consequently, the more general criteria lending themselves to broader analysis are preferred, such as those listed previously.

Typically, these criteria are selected according to their potential for detecting occurrences of fraud or tracking the specific behavior of an entity under analysis, usually by the standards for a particular industry as established by past experience, or by an expert in intelligence gathering and analysis. In essence, the selection of these criteria acts as one step in the filtering process whereby essential data from the large amounts of client transactional data is identified for use in the present invention. This criteria selection is analogous to a method used in military intelligence known as "indications and warning" that monitored a variety of criteria for indications of increasing threat. When such a threat was detected, an alarm was triggered to warn the analyst. This same technique is used in the preferred embodiment to scan through the behavior history of all providers or patients and identify those having the highest risk for fraud or abuse.

In the preferred embodiment, these analysis criteria are pre-programmed into the system. The analysis

criteria is used to identify the data elements contained in the data derived from the transaction process data structure received from the client data source 12. The selected data elements are included in a periodic extract process (e.g., monthly) which produces an extract data file to be processed by the pre-processor 110.

As shown in Fig. 1, once the criteria are established and programmed into the criteria memory 16, the client data converter 14 must then convert the client transaction data received from the client data source 12 for use in the pre-processor 110. Specifically, the client data converter 14 may be implemented as a hardware device or software program (such as a commercial database engine) which accepts the extract data file and transforms it physically and organizationally in preparation for its use by the pre-processor 110.

The extract data file, typically an American Standard Computer Information Interface (ASCII) file or fixed record length file, is converted to a pre-processor input data file using the data structure required by the pre-processor 110. Some of the data elements received from the client data source 12 are directly used as criteria, others must be converted before they are used. For instance, in the preferred embodiment, a criterion might be the number of claims submitted as a direct input data set. Another criterion might also use the ratio of claims per patient as an input set after dividing the number of claims submitted by the number of patients treated.

In addition to converting the data physically, the client data converter 14 also organizes the data in the

order expected by the pre-processor 110. Since peer-level comparison is one of the concepts used by the pre-processor 110, the data must be arranged into groups of like entities. In the preferred embodiment, this would mean sorting the data into physician specialty groups, in the case of banking, this would involve sorting by like account/branch combinations. The objective is to put the data into groups of entities that can be expected to exhibit common behavior based upon their functional characteristics (e.g., all cardiologists treat heart disease).

The data elements selected according to the analysis criteria can be arranged in any order except for the master criteria. The master criteria is the criteria that will have the highest consistency for reflecting an entity's performance. For instance, good master criteria for health care would be either the number of claims submitted, or the number of patients treated -- since either is required for the physician to have any activity during the month. The number of referrals, for instance, would be a poor master criteria since referrals are not required for the physician to operate.

Within each group, entity data is further organized according to time units. The pre-processor 110 can accept various time units (e.g., days, months, years, etc.) based upon the operational characteristics of the domain and the dynamics of the fraud threat. In the preferred embodiment, it is usually adequate to analyze a one year historical sample of data composed of twelve monthly time units. In the case of bank fraud, since the dynamics are much

greater, the system would analyze ninety, one day time units.

The organization of the pre-processor input data file can be theoretically envisioned as a three-dimensional cube or array (as shown in Fig. 8,) where the x-axis contains the entities under analysis, the y-axis contains the criteria that are to be measured in determining the entity's performance, and the z-axis represents time divided into time units. Accordingly, each entity within a group can be evaluated according to each criterion over a specified amount of time. This creates the concept of an entity-criterion "bucket," wherein statistical scores or values can be generated for each entity according to each criterion.

Once the data are in a form capable of being stored in an electronic database 18, they can be connected to an appropriate controller 10 which is adapted to coordinate, control, monitor and execute commands necessary for operating the client data input apparatus 100, the pre-processor apparatus 110, as well as the monitoring and tracking apparatus 120, as shown in Figs. 1, 2 and 3, respectively. The functions performed by controller 10 are communicated to the hardware elements of Figs. 1, 2 and 3 via various control lines running from the controller 10 to its associated equipment of Figs. 1, 2 and 3. The controller 10 can comprise any appropriate microprocessor or minicomputer control device including, but not limited to, a PC-based controller or a dedicated control machine. In addition, the controller can comprise software implemented as a control program.

As shown in Fig. 2, once the pre-processor input data file is constructed and stored in the client database 18, it can be fed into the statistical analysis engine 50 of the pre-processor apparatus 110. The input file is fed to the statistical analysis engine 50 as demonstrated by the following sequence: group 1, criterion 1, entity 1...n; criterion 2, entity 1...n; criterion n, entity 1...n; group 2...n, etc.. An analysis is performed on each entity-criterion combination by determining three types of statistical measures: Value (v), Normalcy (n) and Change (d). These measures are a key component in the reliability and accuracy of identifying fraudulent behavior. They are designed to detect critical characteristics related to fraudulent behavior.

The first statistical measure is generated by the value processor 22. The value processor 22 generates a Value (v) which determines "how important" the entity is within the group. This is a sort of "damage potential" measurement where the entity could be highly suspect in behavior but have an impact in dollars so small that he or she is not worth going after. The measure is simply the average of the total (t) observations for an entity, divided by the average of the total observations for the criterion among all the entities in the group, times one hundred. That is, calculate the mean for the criterion values for an entity by: (1) determining the number of active time units (Nt); (2) finding the sum of the values for all active time units (Sty); and (3) dividing Sty by Nt to find the mean of all time units for the entity (Mty).

When the Mty for each entity is calculated these values are summed for the group (S(Mty)) and then divided by the number of entities in the group (Ne) to find the mean criteria value for the group (Mc). Each entity's Mty is then divided by Mc, and the result is multiplied by 100 to find the (v). The weight (e.g., multiplier) for the value measure for this criteria is looked up in the weight table 46 and then applied to (v) to find the weighted value score (sv). This (sv) is stored in the statistical value memory 28. (Figs. 4 and 4A give detailed block flow diagrams of the processes by which the weighted value (sv) score is generated.)

The second statistical measure is generated by the normalcy processor 24. Normalcy (n) is a measure of how much the entity conforms to the normal behavior of the group for this particular criterion. The measure is important because it is rarely possible that any particular entity can know how his or her behavior appears within the behavior of the total group.

The normalcy (n) measure is the statistical "standard score" for the entity. Calculating a standard score is relatively simple and well known in the art, and therefore any routine software implementation will suffice. Once the standard score for the entity is calculated, the weight for the normalcy measure for this criterion is looked up in the weight table 46 and then applied to (n) to find the weighted normalcy score (sn). This (sn) is stored in the statistical value memory 28. (Fig. 5 gives a detailed block flow diagram of the process by which the weighted normalcy (sn) score is generated.)

The third statistical measures is generated by the change processor 26. Change (d) is measured by performing a standard single linear regression on the observation (t) values for the entity, estimating the value for the last t, comparing the estimated to the actual value for the last t, multiplying the difference by a constant and determining if the actual value is greater or less than the estimate by more than the allowed amount. As with the generation of the Normalcy (n) measure, the calculation is relatively simple and well known in the art, and therefore any routine software implementation for a standard single linear regression will suffice. Once the standard single linear regression for the entity is calculated, the weight for the change measure for this criteria is looked up in the weight table 46 and then applied to (d) to find the weighted change score (sd). This (sd) is stored in the statistical value memory 28. (Fig. 6 gives a detailed block flow diagram of the process by which the weighted change (sd) score is generated.)

Returning to Fig. 2, the entity-criterion score generator 30 generates a single score for each entity-criterion. The entity-criterion score generator 30 retrieves the (sv), (sn) and (sd) scores for each entity-criterion and sums them to form a single entity-criterion score. The weight for each entity-criterion score for each entity is looked up in the weight table 46 and is then applied to the entity-criterion score to find the weighted entity-criterion score which is stored in the entity-criterion memory 32.

The entity score generator 34 takes each of the weighted entity-criterion scores from the entity-criterion

memory 32 and sums all the weighted entity-criterion scores to form a single entity score. This score is then stored in the entity score database 36.

The risk analysis processor 38 takes the single entity score for all entities processed and compares them to a threshold value derived from the threshold table 48. The entities are then given a label corresponding to their amount of risk in relation to the threshold value.

As shown in Fig. 3, the entities are displayed by the risklist display 40, which uses text, tables, graphs, and graphical indicators to identify the levels of risk for each entity with respect to their engagement in fraudulent activities. In the preferred embodiment, entities are identified with a "red ball" to indicate high risk, a "yellow ball" to indicate medium risk, and a "green ball" to indicate low risk. This risklist of entities may be provided to an appropriate printer/plotter device 54 or display 52 to show the degree of risk for each entity.

One object of the present invention is to provide information to a human analyst in order to assist the analyst in making decisions and taking actions on appropriate risk entity targets. The risk action generator 42 offers a list of actions that can be taken against an entity. In the preferred embodiment, these actions can include a notice to the provider of performance aberration, or a request for an explanation of the behavior in a particular area. Actions are carefully defined to be implementable in enterprise operations and traceable within the system. A generic list of actions would include: monitor, notify, warn, threaten, and cancel.

The analyst user of the fraud pre-processor detection system scans the results of the analysis from the risklist display 40 and selects entities that appear to exhibit the highest risk of fraud. The analyst decides whether an action should be taken against an entity. The purpose of taking action is twofold: (1) to induce a change in the entity's behavior that is observable by the system and which will indicate his or her prior fraudulent actions (by highlighting those behavior changes in subsequent analysis); and (2) to modify charges in such a way that savings produced by the fraud pre-processor detection system can be detected and commented upon. The system can also provide support to the conduct of investigations as a result of the observations and conclusions that are developed.

If an action is selected by the analyst user, the actions taken against a particular entity are recorded and monitored by the risk behavior monitor 44. The risk behavior monitor 44 stores the record of the actions, annotations that may be recorded by the analyst, the identity of the analyst, and other information important to maintaining knowledge of the actions related to each entity in the entity status database 47.

The summary and savings generator 45 provides automatic review of the results of the system on a monthly, or other selected period, basis. The summary and savings generator 45 summarizes the number of entities reviewed in each category along with the total amount of dollars involved. Entities identified in various levels of risk are summarized, along with the number and type of actions taken. The system tracks actions and measures taken prior

to action having been taken, and subsequent performance for entities after actions have been taken. The summary and savings generator 45 provides a unique apparatus and process for determining the two most critical factors in anti-fraud management: (1) specific aberrant behavior detection; and (2) cost reduction/savings.

The summary and savings generator 45 is connected to a group summary processor 49 and an individual summary processor 51. The group summary processor 49 performs a monthly summary of the statistics relating to the group under review, such as the size of the group, the numer of high risk entities, amount charged by the group, amount charged by the high risk entities in dollars and percent of the total charges, total actions taken by type (e.g., level 1, 2, 3, etc.), the total amount claimed for savings, and so forth.

The individual summary processor 51 works only on those entities that have been identified with an action flag in the entity status database 47 through the risk action monitor 44 in order to perform specific aberrant behavior detection. When the user analyst initiates one of the actions, the level of the action, the date taken, the date to initiate subsequent review and the evaluation period to review, are all recorded in the entity status database 47.

The individual summary processor 51 operates on a general assumption that most individuals who are knowingly doing something wrong, will, when attention is called to this fact by an external source, modify their behavior to reduce their risk of discovery and/or punitive action.

Consequently, the apparatus and process used in the invention focuses on behavior modification and the subsequent detection of the changes making the assumption that there is a high degree of correlation between the observed changes and the areas of the entity's conduct that were either fraudulent or, at least, suspect.

Individual entities are screened and scored for their "risk" regarding fraud potentials during an initial pass, or first review pass, through the system. The user analyst performs a review of the output scores in the measured criteria generated by the risklist display 40 and determines that some action generated by the risk action generator 42 should be taken against an entity. He uses the risk action monitor 44 to record the action taken. The entity's record is attached with the appropriate action flag, along with the dates of action and the dates for subsequent review and added to the entity status database 47. The normal period of delay from action to subsequent review is 90 days.

The actions generated by the risk action generator 42 are generally a range of options such as: (1) monitor (i.e., place behavior under continuing review without notifying the individual); (2) notify (i.e., communicate with the individual that general behavior patterns have been detected that appear to be in question - but provide only limited details); (3) warn (i.e., communicate to the individual that specific patterns of questionable behavior have been determined, indicating the nature of the behavior and suggesting self review); (4) threaten investigation (i.e., communicate that unacceptable behavior has been detected and that unless corrected

immediately detailed investigation will result); (5) initiate investigation (i.e., communicate that unacceptable behavior of such dimension has been detected that an investigation is being opened and the individual can expect additional action within a specified period of time); and (6) place on probation (i.e., unless detected unacceptable behavior is changed within a specified time, the individual will be dropped as a provider of services/products).

While the exact wording or communications will be unique to each client application, this generic range of actions will be used and the risk action monitor 44 will record each action implemented in the entity status database 47. Since the actions are coded and graded for increasing degrees of severity, the subsequent monitoring of individual performance will take into account the severity of the action when scoring response behavior. For example, a large response (i.e., change in behavior) from a relatively low level action (e.g., #2 - Notify) would be scored higher than a low response for the same action, or for a higher severity action.

The individual summary processor 51 determines which entities to review from the entity status database 47. It also determines from the date of the action and the delay period when a second pass review of the entities should be performed. Once an entity is selected for a second pass review, the individual summary processor 51 initiates the system to analyze client transaction data from the client data source 12 from the date the action was taken to the date when the second pass review was ordered, for that particular entity. The individual summary processor 51 automatically retrieves the prior data on the

entity saved from the first pass review along with the results of the second pass review. The separate passes are extracted from the 36 entity score database as distinct performance periods on either side of the action-taken date. The data for the initial performance period and the subsequent performance period are compared for differences. Downward changes for charges are recorded as potential savings resulting from the action taken. The changes in specific criteria of measurement (e.g., in the preferred embodiment, the number of patients seen; the number of claims submitted; particular procedures rendered) are detected and used as an indicator of potential areas of fraud in the initial (i.e., first pass review) behavior period. Since the severity of the action taken will reflect on the degree of change to be expected, a multiplier from the 46 weight table is applied to amplify the differences observed according to the action taken. Also, since the prior performance of the entity with respect to his or her change trend will bear on the validity of the difference findings, a trend factor from the trend factor table 53 is applied. If the entity's prior performance is highly variable, the validity of the difference findings is reduced. Conversely, if the prior behavior trend is found to be very stable, significant differences will be amplified.

The results of the behavior change analysis performed by the individual summary processor 51 are presented to the analyst. Where the difference results indicate a high degree of correlation with the risk criteria identified in the first review pass, the analyst can assume a causal relationship with the action and that the areas that have changed are suggestive of aberrant

behavior prior to the action. The analyst can then use the behavior changes as a means of reviewing highly specific areas of the entity's prior performance, as well as modify the weight and threshold values found in the 46 weight table and 48 threshold table to better refine the detection process for subsequent reviews. The technique embodied in the invention provides several unique and previously unavailable capabilities to the process of fraud detection.

As shown in Figs. 4, 4A, 5 and 6, details of the block flow diagrams showing the steps for generating the statistical measures for Value (v), Normalcy (n) and Change (d) are shown. As the means for generating these statistical measures are well-known in the art, they will not be described in detail herein. It should be noted, however, that although the means for generating these statistical measures are well-known, the application of these statistical measures to transactional data in order to analyze such data for purposes of fraud detection is considered a novel feature of the invention.

As shown in Fig. 7, which details a block flow diagram showing the steps for performing the pre-processor method, the initial stages of the process serve to filter and highlight suspect individuals, as described previously. Once suspect individuals are detected, prioritized and highlighted by the system an analyst will review the results and actions generated by the risklist display 40 and risk action generator 42, respectively, to determine what action to take. This automated filtering and prioritization process is critical because reviewing all of the behavior of all entities in sufficient detail to precisely identify fraud is impractical. The amount of

data normally collected in the transaction-billing process is far to great to permit manual, or human review. And the potential variations for fraud in such industries as health care are far too many, and much too flexible, to build a completely automated pattern detection system that would be reliable for any protracted period. Consequently, the present invention reduces the focus of the human analyst to the most likely payoff.

The combined statistical approach of the pre-processor apparatus 110 highlights the most critical areas of behavior in performing the filtering, that is: value, normalcy, and change. These measures, when applied to carefully selected behavior criteria (i.e., selected for their potential to exhibit profit-taking) produce a high degree of validity in showing the analyst where to look. The output of the first pass review is normally sufficient for the analyst to decide to take action.

This element of the present invention is one of the unique components of the process - the concept of introducing process control theory to the detection of fraud in large transaction data histories. There has not, heretofore, been available a methodology that was implementable in an automated process that could record behavior changes in specific entities and determine the savings resulting from the counter-fraud program. As a result of the first pass review, organizations currently suffering from high losses due to fraud have a means of not only identifying candidate targets, but they can modify the behavior of those targets in order to obtain savings and further focus on specific fraudulent behavior. The results of the process can offer the analyst with the option of

declaring savings or using the highlighted areas of change as a means of initiating deeper investigation into very specific areas of the entity's performance.

While the initial use of the process is focused on fraud in the health care industry, it is apparent that the technique described here can also be used to detect fraud and identify resulting savings in a variety of environments (e.g., banking, mortgage, worker's compensation, etc.).

I CLAIM:

1. A fraud detection system for targeting a potentially fraudulent entity within a particular industry, comprising:

statistical analysis engine for generating statistical values for said entity according to at least one of a plurality of analysis criteria;

entity-criterion score generator for weighting and summing each of said statistical values thereby forming an entity-criterion score for each of said analysis criteria;

entity score generator for weighting and summing each said entity-criterion scores to form a first entity score;

computer memory for storing said statistical values, said entity-criterion scores, and said first entity score; and

risk analysis processor for comparing said first entity score to a predetermined threshold, whereby the result of said comparison indicates whether said entity is engaging in fraudulent activity.

2. The system according to claim 1, further comprising controller means for coordinating, controlling, monitoring and executing commands necessary for operating said fraud detection system.

3. The system according to claim 1, further comprising a database for storing said first entity score.

4. The system according to claim 1, wherein said statistical values include a value statistic (v), an normalcy statistic (n), and a change statistic (d) for each of said criteria.

5. The system according to claim 1, further comprising client data converter for converting client records containing data elements useful in measuring the performance of said entity to a format compatible with said fraud detection system.

6. The system according to claim 1, further comprising a three dimensional array for organizing and storing said entity and said criteria according to predefined time units, said array comprising an x, y and z axis, said y-axis containing said entity, said x-axis containing said criteria, and said z-axis containing said time units.

7. The system according to claim 6, wherein said array creates entity-criterion buckets wherein said statistical values for each criterion can be generated for said entity within said time period.

8. The system according to claim 1, further comprising a risklist display for displaying in graphical form whether said entity is engaging in fraudulent behavior.

9. The system according to claim 1, further comprising:

risk action generator for displaying possible actions to initiate against said entity, and selecting one of said actions to modify said entity's behavior;

risk behavior monitor for storing a record of said selected action taken against said entity; and

summary and savings generator for identifying the amount of money gained or lost by said action.

10. The system according to claim 9, wherein said summary and savings means include:

a second entity score presenting said entity's behavior after said action has occurred;

individual summary processor for comparing said first entity score with said second entity score, measuring the degree of change therefrom, weighting the degree of change using a severity factor and a trend factor, and displaying a result indicating whether said action decreased the fraudulent activity of said entity;

group summary processor for generating group statistics for a plurality of entities.

11. The system according to claim 10, wherein said individual summary processor includes adjuster for adjusting said weighting values for said criteria scoring means and said entity scoring means, and said threshold

values for said risk analysis means, in accordance with said group statistics from said summary and savings means.

12. A fraud detection system for targeting potentially fraudulent entities within a particular industry, comprising:

client data source for in putting client records containing data elements useful in measuring the performance of an entity by evaluating a plurality of criteria over a set of time period;

there dimensional array for organizing and storing said entity, said criteria, and said time period, comprising an x, y and z axis, said y-axis containing said entity, said x-axis containing said criteria, and said z-axis containing at least one of said time units;

statistical analysis engine for generating a value statistic (v), a normalcy statistic (n), and a change statistic (d) for each of said criteria;

entity-criterion score generator for weighting and summing each of said statistical values thereby forming an entity-criterion score for each of said criteria;

entity score generator for weighting and summing each of said entity-criterion scores to form an entity score for said entity;

risk analysis processor for comparing said entity score with pre-defined threshold values; and

risklist display for displaying in a graphical manner whether said entity is engaging in fraudulent behavior.

13. A computer program for targeting potentially fraudulent entities within a particular industry, said computer program comprising the steps of:

inputting client records containing data elements useful in measuring the performance of an entity by evaluating a plurality of criteria over a set time period;

organizing and storing said entity, said criteria, and said time period, into a three dimensional array comprising an x, y and z axis, said y-axis containing said entity, said x-axis containing said criteria, and said z-axis containing at least one of said time units;

generating a value statistic (v), a normalcy statistic (n), and a change statistic (d) for each of said criteria;

weighting and summing each of said statistical values thereby forming an entity-criterion score for each of said criteria;

weighting and summing each of said entity-criterion scores to form an entity score for said entity;

comparing said entity score with pre-defined threshold values; and

displaying graphically whether said entity is engaging in fraudulent behavior.

14. The computer program according to claim 13, said computer program further comprising the step of encoding said computer program on a computer-readable medium.

15. A method of targeting potentially fraudulent entities within a particular industry, comprising the steps of:

generating statistical values for entity according to a plurality of analysis criteria;

weighting and summing each of said statistical values thereby forming an entity-criterion score for each of said analysis criteria;

weighing and summing each of said entity-criterion scores to form an entity score; and

comparing said entity score to a predetermined threshold.

16. The computer program according to claim 15, said computer program further comprising the step of encoding said computer program on a computer-readable medium.

17. A fraud pre-processor system for filtering and highlighting those individuals most likely to be engaging in fraudulent behavior, comprising:

a controller;

a computer program executed by said controller comprising:

client data converter for converting client data to format compatible with said fraud pre-processor system;

three dimensional array for storing a plurality of entities, analysis criteria, and time units such that for each of said entities statistical values for each of said criteria within each of said time units can be generated, weighted and summed to form an entity-criterion score;

entity-criterion score generator for generating said entity-criterion score;

entity score generator for weighting and summing each entity-criterion score for each entity to form an entity score;

risk analysis processor for comparing said entity score for each of said entities with a threshold value;

risklist display for displaying the results from said comparison in a manner identifying those entities most likely engaging in fraudulent behavior; and

a database for permanently storing, and a computer memory for temporarily storing, said analysis criteria, said statistical values, said entity-criterion score, said entity-score, said threshold values, and said results.

18. A fraud detection system, comprising statistical analysis engine for generating statistics for a plurality of entities according to a plurality of analysis criteria and time periods; and

pre-processor for evaluating said statistics to determine if said entities are engaging in fraudulent behavior.

19. The system according to claim 1, wherein said statistical analysis engine further comprises:

value processor for generating said value statistic (v) which determines the damage potential for said entity;

normalcy processor for generating said normalcy statistic (n) for measuring normal behavior for said entities; and change processor for generating said change statistic (d).

20. A fraud detection system, comprising:

risk analysis processor for comparing an entity score for an entity to a predetermined threshold;

risk action generator for displaying and selecting action to be taken said entity;

risk action method for storing a record of said actions; summary and savings generator for identifying the amount of money gained or lost by said action; and

adjustor for adjusting weighting values used to generate said entity score, and said threshold values for said risk analysis means, in accordance with said statistics from said summary and savings generator, in order to better estimate if said entity is engaging in fraudulent behavior.

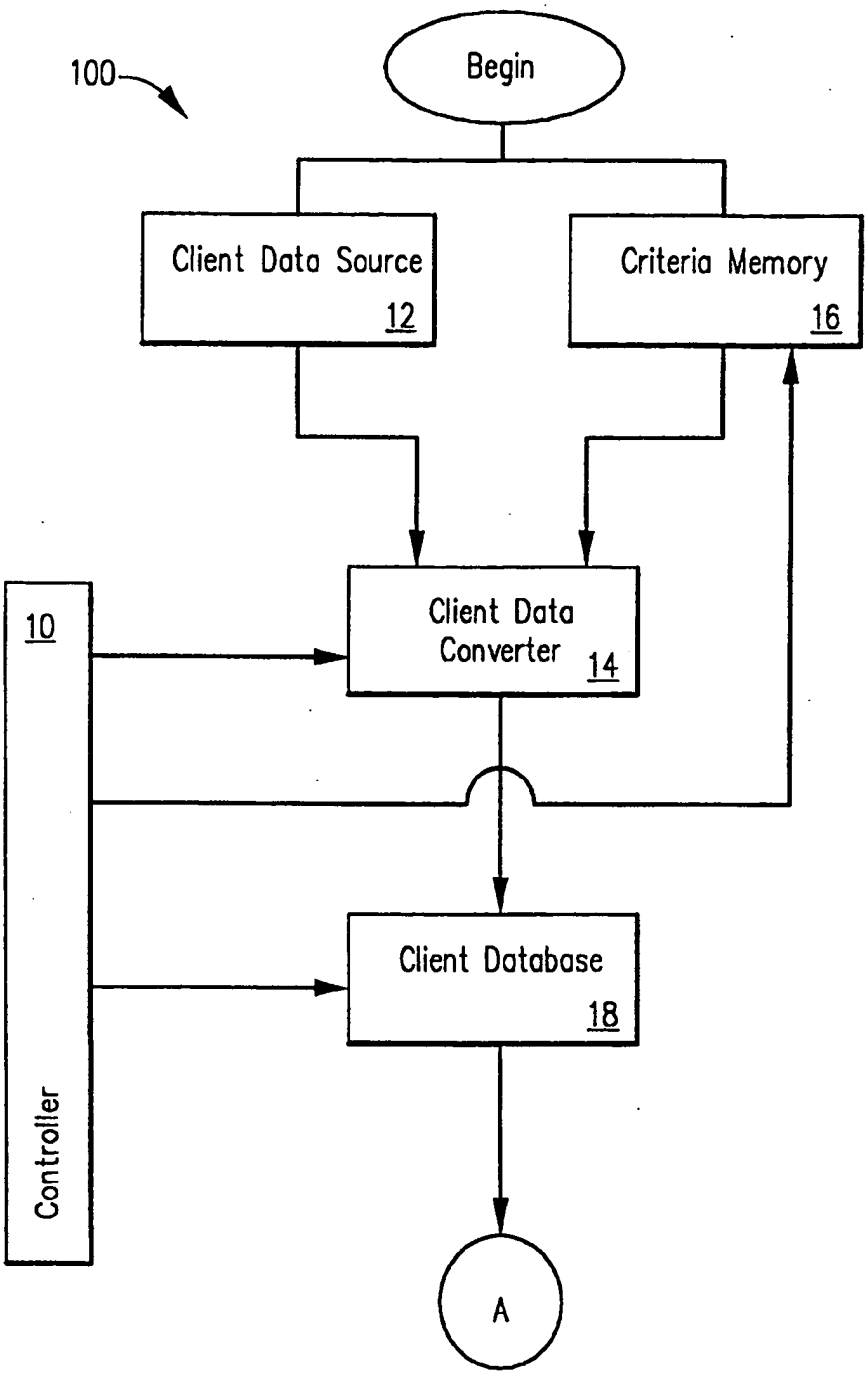

100
Begin
Client Data Source
12
Criteria Memory
16
Client Data
Converter
14
10
Client Database
18
Controller
A

FIG. 1

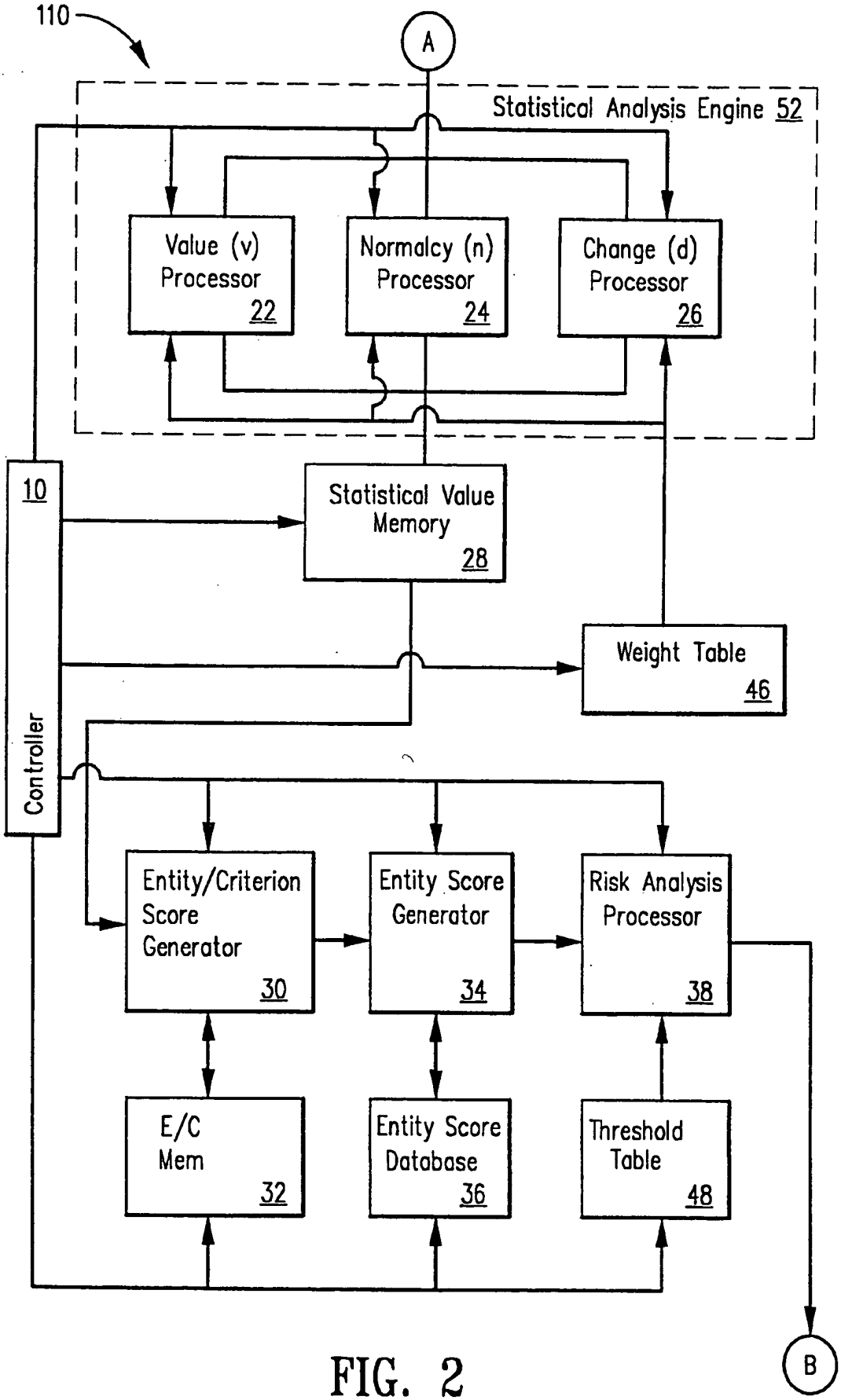

110
A
Statistical Analysis Engine 52
Value (v) Processor 22
Normalcy (n) Processor 24
Change (d) Processor 26
10
Statistical Value Memory 28
Weight Table 46
Controller
Entity/Criterion Score Generator 30
Entity Score Generator 34
Risk Analysis Processor 38
E/C Mem 32
Entity Score Database 36
Threshold Table 48
B

FIG. 2

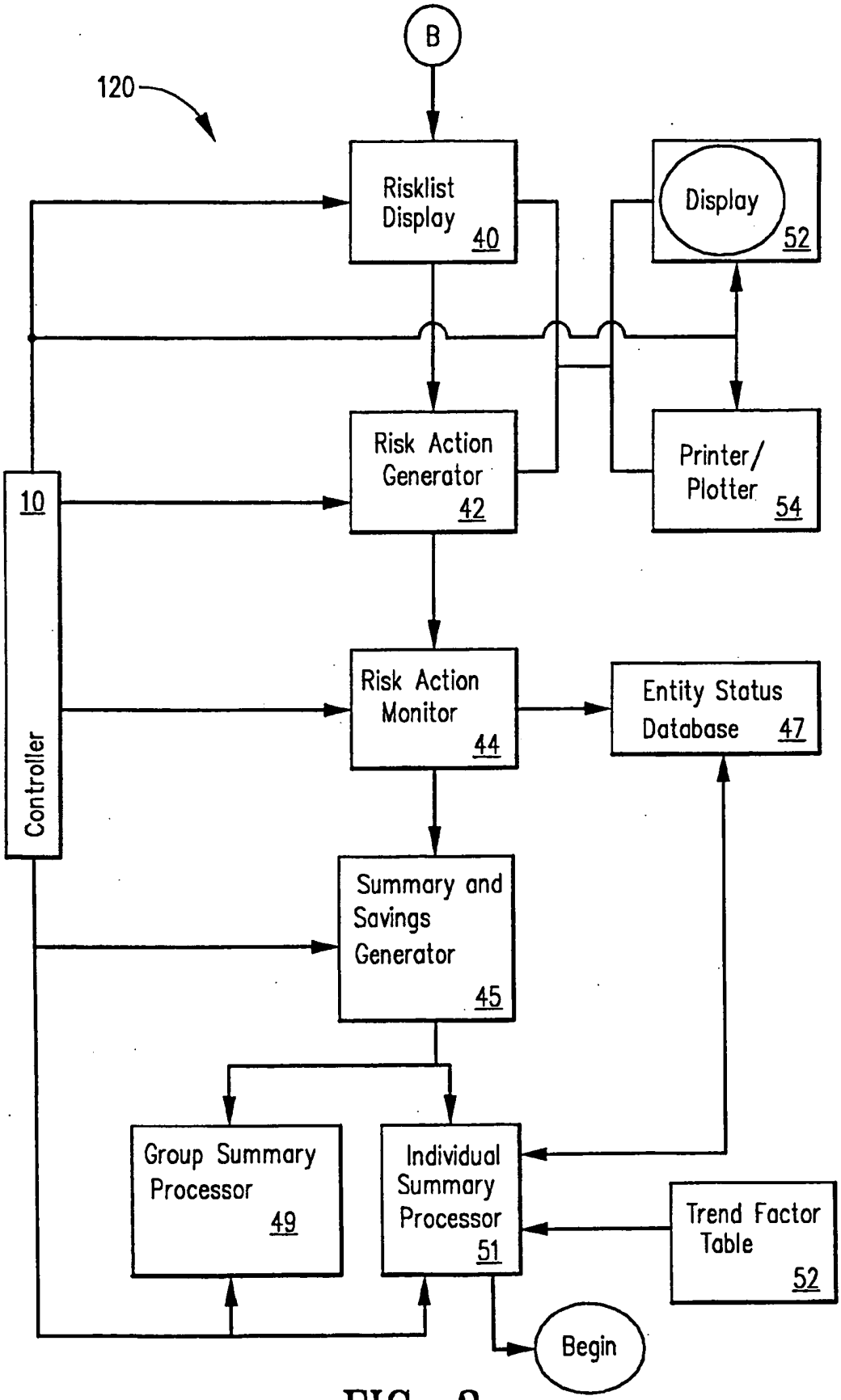

B
120
Risklist Display 40
Display 52
Risk Action Generator 42
Printer/ Plotter 54
10
Controller
Risk Action Monitor 44
Entity Status Database 47
Summary and Savings Generator 45
Group Summary Processor 49
Individual Summary Processor 51
Trend Factor Table 52
Begin

FIG. 3

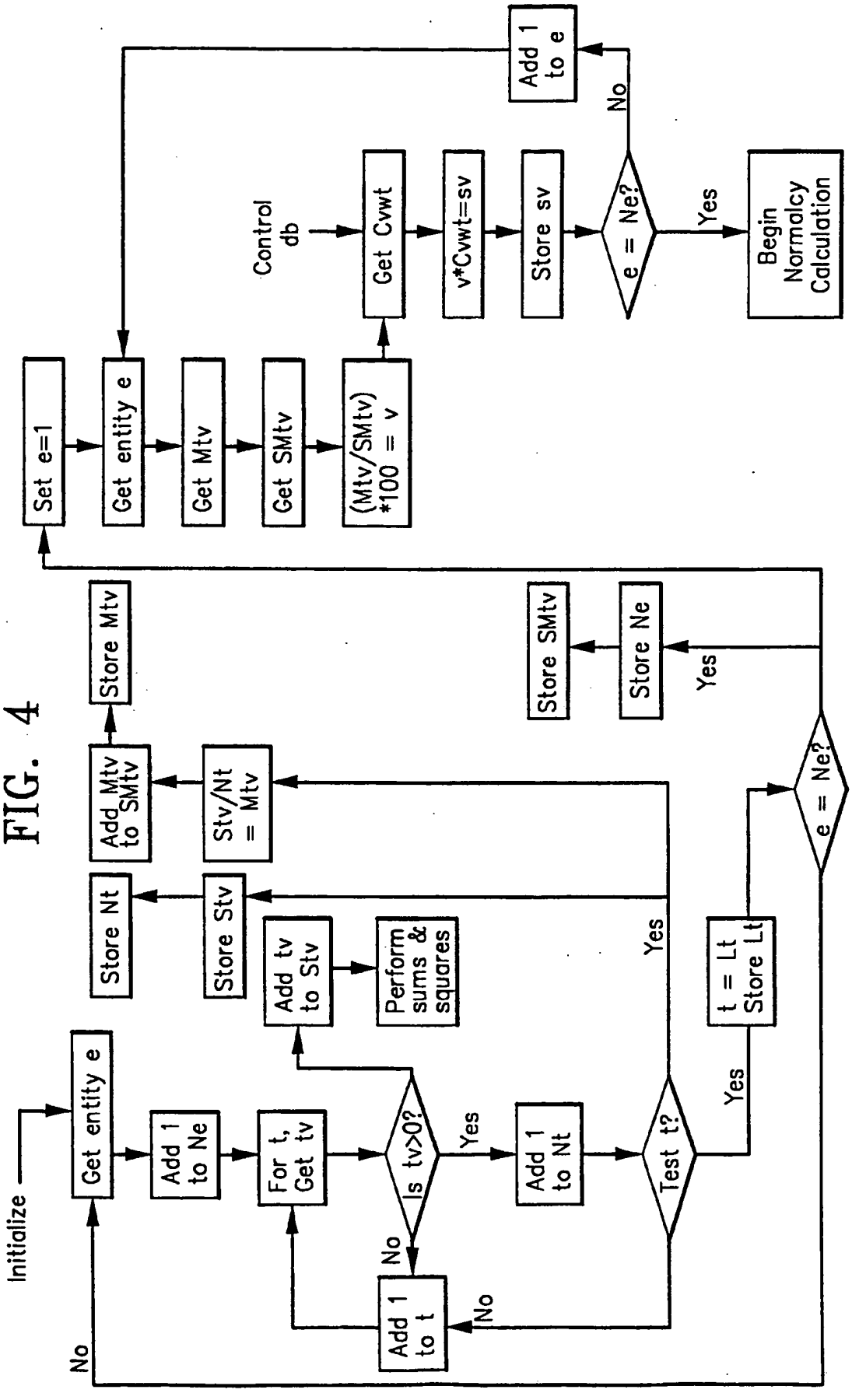

Initialize
Get entity e
Add 1 to Ne
For t, Get tv
Is tv>0?
No
Yes
Add 1 to t
Add 1 to Nt
Test t?
No
Yes
t = Lt Store Lt
Add tv to Stv
Perform sums & squares
Store Nt
Store Stv
Add Mtv to SMtv
Stv/Nt = Mtv
Store Mtv
e = Ne?
No
Yes
Store Ne
Store SMtv
Set e=1
Get entity e
Get Mtv
Get SMtv
(Mtv/SMtv) *100 = v
Control db
Get Cvwt
v*Cvwt=sv
Store sv
e = Ne?
No
Yes
Add 1 to e
Begin Normalcy Calculation

FIG. 4

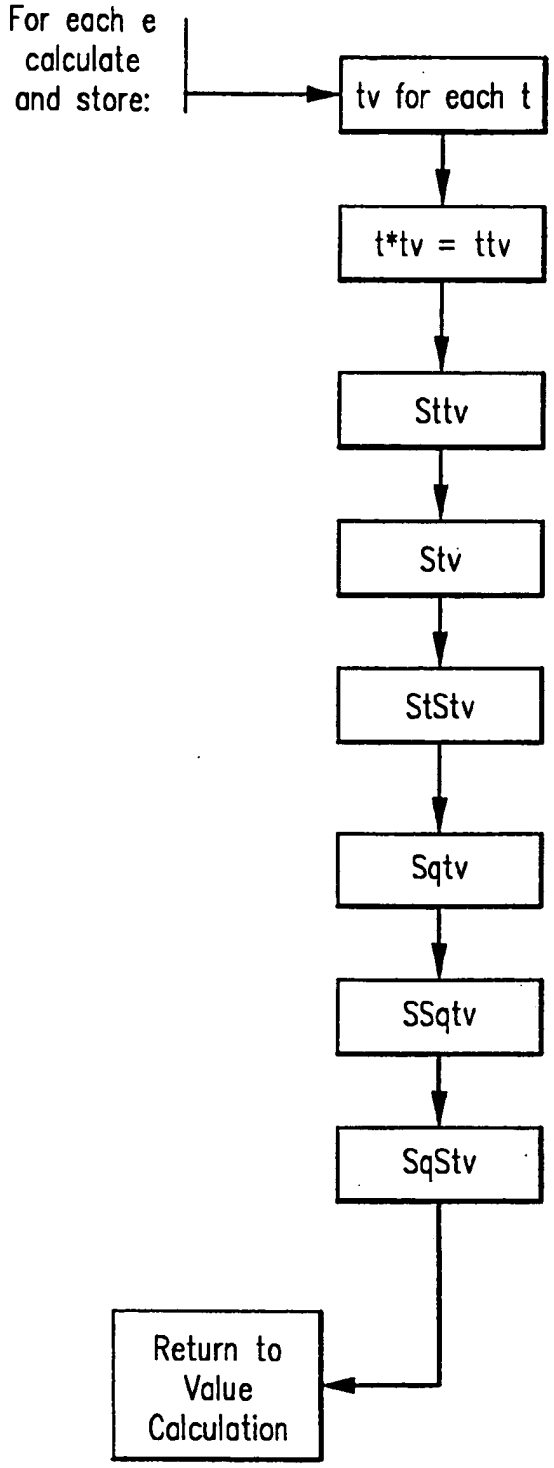

For each e
calculate
and store:
tv for each t
t*tv = ttv
Sttv
Stv
StStv
Sqtv
SSqtv
SqStv
Return to
Value
Calculation

FIG. 4A

FIG. 5

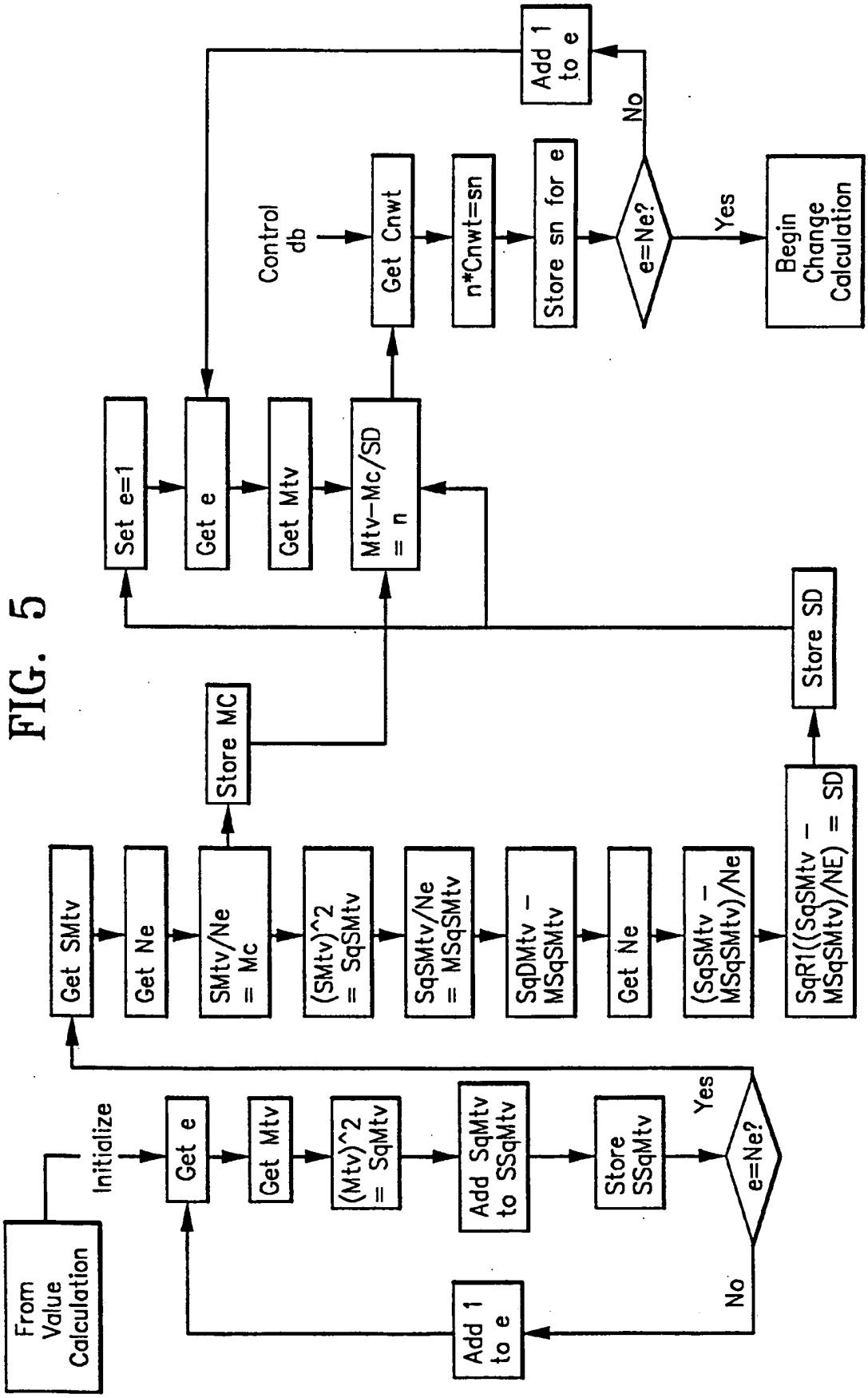
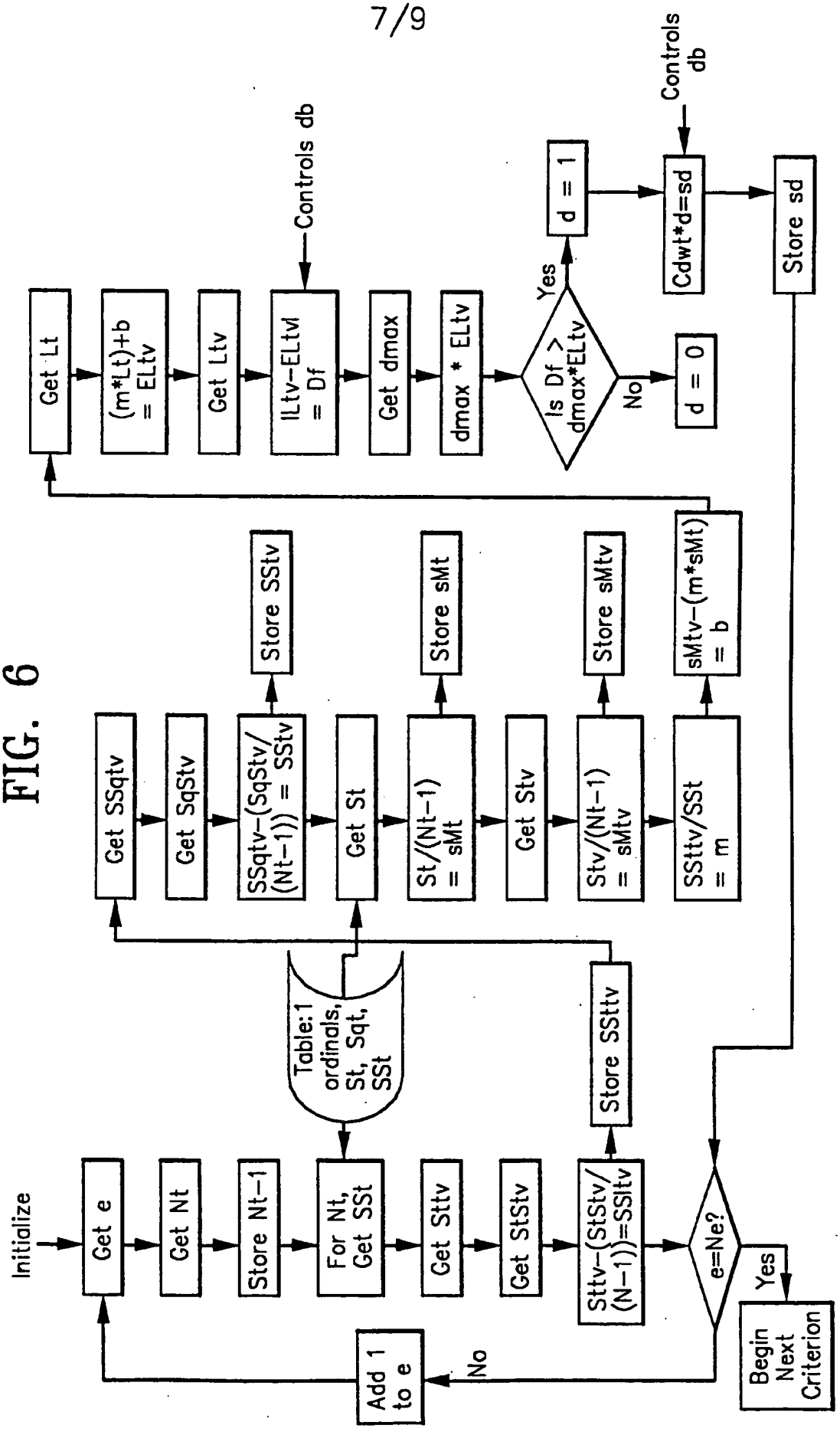

From Value Calculation
Initialize
Get e
Get Mtv
(Mtv)^2 = SqMtv
Add SqMtv to SSqMtv
Store SSqMtv
e=Ne?
Yes
No
Add 1 to e
Get SMtv
Get Ne
SMtv/Ne = Mc
(SMtv)^2 = SqSMtv
SqSMtv/Ne = MSqSMtv
SqDMtv - MSqSMtv
Get Ne
(SqSMtv - MSqSMtv)/Ne
SqR1((SqSMtv - MSqSMtv)/NE) = SD
Store MC
Store SD
Set e=1
Get e
Get Mtv
Mtv-Mc/SD = n
Control db
Get Cnwt
n*Cnwt=sn
Store sn for e
e=Ne?
Yes
No
Add 1 to e
Begin Change Calculation

FIG. 6


Initialize
Get e
Get Nt
Store Nt-1
For Nt, Get SSt
Get Sttv
Get StStv
Sttv-(StStv/(N-1))=SSttv
Store SSttv
e=Ne?
Yes
No
Add 1 to e
Begin Next Criterion
Table: 1 ordinals, St, Sqt, SSt
Get SSqtv
Get SqStv
SSqtv-(SqStv/(Nt-1)) = SStv
Store SStv
Get St
St/(Nt-1) = sMt
Store sMt
Get Stv
Stv/(Nt-1) = sMtv
Store sMtv
SSttv/SSt = m
sMtv-(m*sMt) = b
Get Lt
(m*Lt)+b = ELtv
Get Ltv
ILtv-ELtvI = Df
Controls db
Get dmax
dmax * ELtv
Is Df > dmax*ELtv
Yes
No
d = 1
d = 0
Controls db
Cdwt*d=sd
Store sd

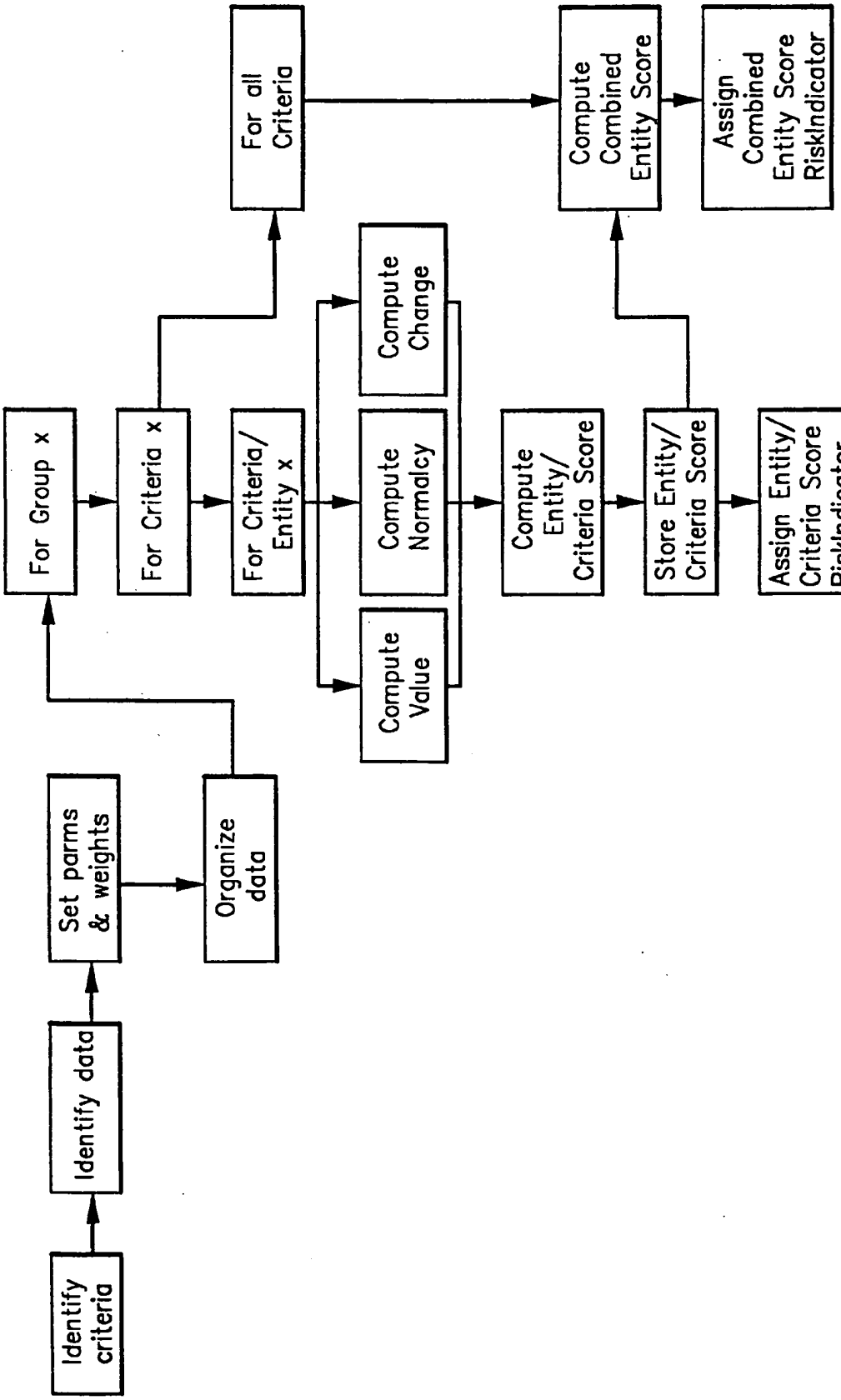

Identify criteria
Identify data
Set parms & weights
Organize data
For Group x
For Criteria x
For Criteria/ Entity x
Compute Value
Compute Normalcy
Compute Change
Compute Entity/ Criteria Score
Store Entity/ Criteria Score
Assign Entity/ Criteria Score RiskIndicator
For all Criteria
Compute Combined Entity Score
Assign Combined Entity Score RiskIndicator

FIG. 7

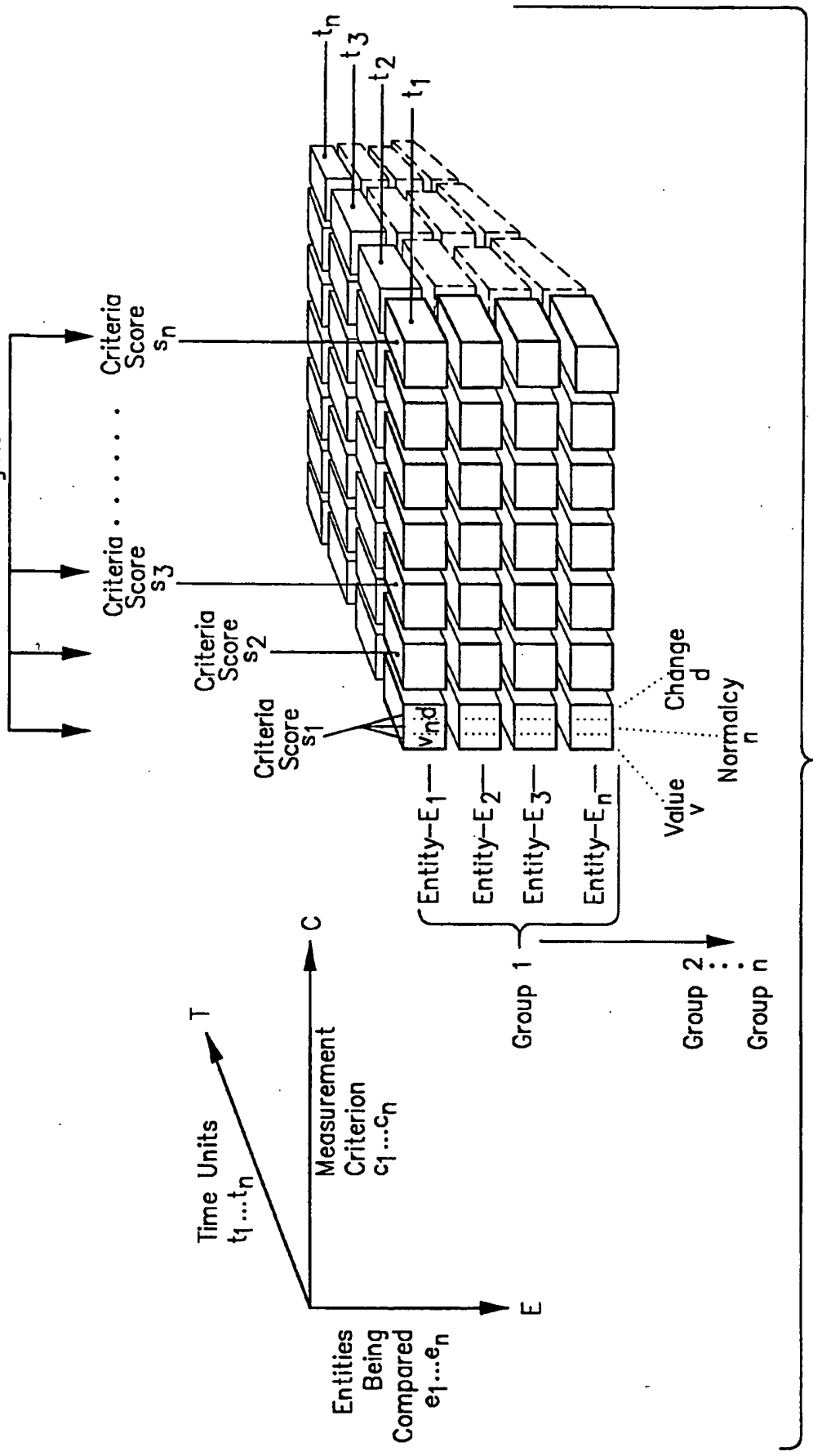

Criterion Weights
Criteria Score s_n
Criteria Score s_3
Criteria Score s_2
Criteria Score s_1
t_n
t_3
t_2
t_1
v_n
d
Change d
Normalcy n
Value v
Entity-E_1
Entity-E_2
Entity-E_3
Entity-E_n
Group 1
Group 2
Group n
T
Time Units $t_1 \ldots t_n$
C
Measurement Criterion $c_1 \ldots c_n$
E
Entities Being Compared $e_1 \ldots e_n$

FIG. 8